بسم الله الرحمن الرحيم

مدخل إلى نظرية القياس
التقليدية والمعاصرة

عنوان الكتاب: مدخل إلى نظرية القياس التقليدية والمعاصرة
تأليف: ليندا كروكر و جيمز الجينا / ترجمة: د. زينات يوسف دعنا
رقم التصنيف: **370.15**
رقم الايداع لدى دائرة المكتبة الوطنية: 3244/09/2008
الموضوع الرئيسي: علم نفس تربوي//قياس التحصيل//مؤشرات الانجاز
تم إعداد بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية

الطبعة الأولى، 2009 - 1430
حقوق الطبع محفوظة

دار الفكر
ناشرون وموزعون

www.daralfiker.com
المملكة الأردنية الهاشمية - عمّان
ساحة الجامع الحسيني - سوق البتراء - عمارة الحجيري
هاتف: 4621938 6 962+ فاكس: 4654761 6 962+
ص.ب: 183520 عمان 11118 الأردن
بريد الكتروني: info@daralfiker.com
بريد المبيعات: sales@daralfiker.com

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means without prior permission in writing of the publisher.

جميع الحقوق محفوظة. لا يسمح بإعادة إصدار هذا الكتاب أو أيّ جزء منه، أو تخزينه في نطاق استعادة المعلومات، أو نقله بأيّ شكل من الأشكال، دون إذن مسبق من الناشر.

ISBN: 9957-07-608-5

مدخل إلى

نظرية القياس التقليدية والمعاصرة

ليندا كروكر   جيمز الجينا

ترجمة

د. زينات يوسف دعنا

الطبعة الأولى

1430-2009

دار الفكر

ناشرون وموزعون

مقدمة المؤلفان

نشر ثورندايك عام 1904 مؤلفه مدخل إلى نظرية القياس العقلي والاجتماعي والذي يعد أول كتاب منهجي في نظرية القياس. وقد رحب بهذا العمل بحماس كبير من قبل زملائه والمختصين في المجال نفسه. وبعدها وخلال فترة وجيزة أصبحت دراسة نظرية القياس مقرر أساسي من منهج الدراسات العليا في كل من التربية وعلم النفس. وعبر العقود الستة التالية اتسعت نجاحات العلماء ونقحت وأضيفت إلى هيكل النظرية التي وصفها ثورندايك. وهذا الهيكل المجتمع من المعرفة عرف باسم نظرية القياس التقليدية، وهذه قدمت لنا الأساس النظري لتطوير معظم مقاييس الاستعداد والتحصيل والشخصية والميول المستخدمة في القرن العشرين. وتقدمت نظرية القياس في العشرين سنة الأخيرة بشكل سريع جعل من المستحيل تدوين كل تقدمها في الكتب المقررة التقليدية. ويعزى هذا التقدم المعرفي في جزء منه إلى تطور التقنيات الحوسبية التي جعلت من الممكن تطبيق نماذج رياضية وإحصائية أكثر تعقيداً لدرجات الاختبارات. وبازدياد عدد المختصين في القياس السيكومتري والحاجة إلى القياس الموضوعي للطلبة والمسترشدين ثم الاستفادة من ذخيرة الخدمات العامة والبرامج التربوية التي ساهمت في هذا المجال. وتبعا لذلك فان تفحص محتوى مجلات القياس الرائدة في التربية وعلم النفس توحي بأن العديد من الموضوعات السائدة حديثة في أساسها.

ومن الأمثلة على مثل هذه الموضوعات طرائق تطوير اختبارات محكية المرجع، ونظرية الاستجابة للفقرة ونظرية إمكانية التعميم والطرائق السيكومترية المتبعة في الكشف عن تحيز الفقرات الاختبارية. ومن منطلق هذا التطور فإن المعرفة بنظرية القياس التقليدية لم تعد كافية لإعداد طلبة الدراسات العليا أو المختصين في القياس، بل إن الحاجة ملحة للإطلاع على الدراسات الحديثة في هذا المجال وتطبيق المعلومات المتضمنة في هذه الدراسات.

إن طلبة نظرية القياس الحديثة يجب أن يكتسبوا معرفة أساسية بالقياس السيكومتري التقليدي إضافة إلى أنه يجب أن تكون لديهم القدرة على دمج أفكار جديدة إلى هيكل

المعرفة. وقد وضع هذا الكتاب ليساعد القارئ في اكتساب هذه الأمور، والقارئ الذي يأمل أن يجد سلسلة خطوات متتابعة لإجراء أي عملية من خلال النقاش التقني للرموز الاحصائية سيخيب امله. وقد بينا ان الطرائق الافضل او الموصى بها بدرجة اكبر لاي من مظاهر تطوير القياس قد تتغير كلما ظهرت وتأسست افكار ونتائج تجريبية جديدة. لذا على طلبة القياس ان يكتسبوا بعض المراس في قراءة المحتوى الذي يتضمن المصطلحات والرموز التقنية المشابهة لما سيحتاجونه عندما ينخرطون في سلك الدراسات العليا والبدء بالقراءة المتخصصة لادبيات القياس وباستقلالية . وفي الوقت نفسه فنحن مقتنعون بان العديد من القراء ليس لديهم اكثر من مساق مدخلي في الاحصاء. لذا فان التعبيرات الكمية اتبعت بالتفسير اللفظي والذي نأمل ان لا يبدو مسهبا أو مملاً للقارئين ذوي الخلفية الجيدة. وتتطلب المعالجة الدقيقة او المناسبة في بعض الاحيان محتوى يفوق الطرائق الاحصائية المدخلية (على سبيل المثال محتوى كل من نظرية امكانية التعميم ونظرية الاستجابة للفقرة). وفي مثل هذه الحالات فانه من الأهمية بمكان دراسة هذا بدلا من إهمالها أو حذفها حتى تصبح بسيطة، لذا فان القارئين الذين لديهم القليل من الخبرة في هذا الموضوع عليهم استخدام هذه الطرائق او تفسير نتاجاتها ونتائجها.

وقد نظمت المادة التعليمية في هذا الكتاب ضمن خمس وحدات هي: مدخل الى نظرية القياس، والثبات، والصدق، وتحليل الفقرة في تطوير الاختبار، وتصحيح الاختبار وتفسير درجاته، وتعرض الوحدة الاولى معلومات اساسية مع تدريبات بسيطة في القياس والاحصاء. وكل وحدة من الوحدات التالية تعرض طريقة القياس التقليدية في الفصول الاولى. وعرضت معلومات اكثر حداثة في الفصول التالية مع مناقشة لكيفية ربط هذه الطرائق بالمفاهيم التقليدية. عموماً، فان التطورات الحديثة في معظم مجالات نظرية القياس تتطلب معالجات اكثر تعقيدا. لذا فان فصول كل وحدة مرتبة من الاسهل الى الاصعب والاكثر تعقيداً من حيث المحتوى في الموضوع الواحد. وعلى المدرسين والمشرفين الذي يعينون قراءات معينة لطلبتهم ان يتنبهوا الى ان الجزء الاخير في معظم الفصول يمكن حذفها او تعيينها للقراءات الاضافية الاختيارية دون فقدان الاستمرارية مع محتوى الوحدات التالية. وان كانت الرغبة محصورة في نظرية القياس التقليدية فيمكن الحصول عليها في الفصول من الاول حتى السابع والعاشر والرابع عشر والسابع عشر والتاسع

عشر. ويمكن ايجاد مقدمة عن الموضوعات السيكومترية الحديثة ولمن لديهم خلفية عن نظرية القياس التقليدية في الفصول الثامن والتاسع والخامس عشر والسادس عشر والثامن عشر والعشرون. هذا وتعد التمارين الحسابية والاسئلة المبينة في نهاية كل فصل جزء مكمل لهذا الكتاب، فقد طورت خصيصاً لتوضيح تطبيقات المفاهيم الاساسية المبينة في الفصل وتوفر للقارئ فرصة الاستخدام النشط لتطبيق نظرية القياس في مشكلات القياس الحالية، ويوفر الملحق ب اجابات على العديد من التمارين الحسابية.

ان تأليف كتاب منهجي يعد عملاً شاقاً ومثبطاً للعزيمة ومذلاً لكنه يغني مؤلفه بخبرة غنية. ومن مصادر تثبيط العزيمة هو عدم كفاية معرفتنا الخاصة وقدراتنا المحدودة في النواحي المنطقية والحسابية والاخطاء القواعدية في التنظيم والتبسيط والترجمة الدقيقة لافكار الاخرين في عمل بيداغوجي متكامل. وتتأتى مكافئة هذا الجهد من اعطائها الفرصة لدراسة الاعمال الاصيلة للعديد من الزملاء الاساتذة في المجال ومن الدعم المخلص الاصيل من جهات عديدة. ونحن مدينين في مؤلفنا هذا لطلبة نظرية القياس في جامعة فلوريدا الذين شجعوا هذا العمل مع تقديمهم لاقتراحات بناءة لمخطوطة الكتاب في وقت اعدادها وتطويرها بالاضافة الى مساعدتهم في مراجعة واعداد تمارين نهاية الفصول. كذلك تلقينا ملاحظات ومقترحات وزودنا بمراجع مفيدة من قبل العديد من الزملاء المتخصصين الذين قرأوا المخطوطة الاولية وقدموا الملاحظات والمقترحات المفيدة. هذا بالاضافة الى جهود القائمين على طباعة واخراج هذا الكتاب وهم هيئة التحرير في Holt, Rinehart Swinstor وكانت جهودهم كبيرة. ويعد هذا العمل مدعوماً ونتاجاً لمساهمات هؤلاء جميعاً.

المؤلفان

كروكر والجنيا

بسم الله الرحمن الرحيم

مقدمة المترجمة

" الحمد لله الذي هوانا لهذا وما كنا لنهتدي لولا أن هدانا الله " وأصلي وأسلم على سيد المرسلين المبعوث رحمة للعالمين ومن تبعه بإحسان إلى يوم الدين.

فقد منَّ الله عليّ ووفقني لترجمة هذا الكتاب " نظرية القياس التقليدية والمعاصرة "، والذي يعد من المراجع الأساسية في القياس النفسي والتربوي والذي لا يستغني عنه المتخصصين في مجال القياس.

يعرض هذا الكتاب نظريات القياس الأساسية الثلاث: نظرية القياس التقليدية ونظرية التعميم (نظرية اتخاذ القرار) ونظرية السمة الكامنة بنماذجها الأساسية ويتضمن ذا العرض الخلفية النظرية والخطوات المنهجية والاستخدامات العملية والتطبيقات مع أمثلة متنوعة من أدبيات القياس، مع الموازنة بين تطبيقات نماذج هذه النظريات وفق أسس نظرية رياضية وتطبيقية.

وأسأل الله العليّ القدير أن يتقبل هذا العمل المتواضع وأن يكون خالصاً لوجهه الكريم.

والله من وراء القصد

المترجمة

المحتويات

1 الوحدة الاولى: مدخل الى نظرية القياس

الفصل الاول: ما هي نظرية القياس 19
مشكلات في قياس الابنية النفسية 21
نظرية القياس على انها فرع من فروع المعرفة 23
تنظيم الكتاب 29
الخلاصة 30
التمارين 30

الفصل الثاني: المفاهيم الاحصائية لنظرية القياس 36
درجات الاختبار كمتغيرات منقطعة لمجتمعات محددة 36
وصف الاداء الفردي بالدرجات الزائية 42
المجتمعات غير المحددة والمتغيرات المتصلة 44
التوزيع الطبيعي 45
وصف العلاقة بين متغيرين 49
التنبؤ بالاداء الفردي 58
الخلاصة 64
التمارين 65

الفصل الثالث: مدخل الى التدريج 72
الاعداد الحقيقية وتداريج القياس 72
مستويات تدريج القياس 73
اساليب التدريج في تطوير القياس 76
مستويات التدريج للمقاييس المتمركزة حول الفرد 90
الخلاصة 93
التمارين 95

الفصل الرابع:عملية بناء الاختبار 98
تحديد اهداف استخدام درجات الاختبار 99
تحديد السلوكات الممثلة للبناء 99
معاينة المجال 101
اعداد مواصفات الاختبار 105
بناء الفقرات 109
مراجعة الفقرات 116
التجريب الاولي للفقرات 118
الخطوات التالية 119
الخلاصة 119
التمارين 121

الفصل الخامس: درجات الاختبار كدرجات مركبة 126
مخططات تصحيح الفقرات 126
الاحصاء الوصفي للمتغيرات اللاثنائية 128
الاحصاء الوصفي للمتغيرات الثنائية 130
تباين المركب 135
تضمينات عملية في بناء الاختبار 139
الخلاصة 142
التمارين 143

2 الوحدة الثانية: الثبات

الفصل السادس: الثبات ونموذج الدرجة الحقيقية التقليدي 147
نموذج الدرجة الحقيقية التقليدي 148
دليل الثبات ومعامل الثبات 157
ثبات المركب 161
الخطأ المعياري للقياس 167
تعريفات بديلة للدرجات الحقيقية والخطأ 169

الخلاصة ———— 173
التمارين ———— 175

الفصل السابع: طرائق حساب الثبات ———— 180
طرائق تتطلب تطبيقين للاختبار ———— 180
طرائق تتطلب تطبيق واحد للاختبار ———— 184
ثبات تقديرات المقدرين ———— 195
العوامل المؤثرة في معامل الثبات ———— 195
الدرجات الحقيقية ———— 199
ثبات الفروق في الدرجات ———— 201
استخدام تقديرات الخطأ في تفسير الدرجات ———— 204
تدوين بيانات الثبات ———— 206
الخلاصة ———— 207
التمارين ———— 209

الفصل الثامن: مدخل الى نظرية امكانية التعميم ———— 216
دراسات G ودراسات D ———— 217
صياغة معاملات امكانية التعميم للتصميمات احادية البعد ———— 219
دراسة G لبعد واحد متداخل ———— 232
النطاقات والابعاد الثابتة ———— 233
الاخطاء المعيارية للقياس للقرارات المطلقة والنسبية ———— 236
نظرية امكانية التعميم للتصميمات ثنائية البعد ———— 240
الخلاصة ———— 253
التمارين ———— 254

الفصل التاسع: معاملات ثبات الاختبارات محكية المرجع ———— 260
استخدامات القياس محكي المرجع ———— 261
نظرية الثبات لتقدير درجة النطاق ———— 261

نظرية الثبات لتصنيفات السيطرة ---- 266
دقة القرار ---- 281
الخلاصة ---- 283
التمارين ---- 285

3 الوحدة الثالثة: الصدق

الفصل العاشر: مقدمة الى الصدق ---- 291
صدق المحتوى ---- 292
الصدق المرتبط بمحك ---- 299
صدق البناء ---- 307
التداخل عبر طرائق الصدق ---- 313
معاملات صدق الدرجات الحقيقية ---- 314
الخلاصة ---- 316
التمارين ---- 317

الفصل الحادي عشر : الطرائق الاحصائية للتنبؤ والتصنيف ---- 324
الارتباط الجزئي ---- 324
الانحدار المتعدد ---- 328
تحليل التمييز ---- 341
الخلاصة ---- 350
التمارين ---- 351

الفصل الثاني عشر: التحيز في الاختيار ---- 356
المصطلحات والمفاهيم الاساسية ---- 357
المجموعات الرئيسة والفرعية ---- 360
التعريفات النفسية للتحيز ---- 360
طرائق الاختيار العادلة ---- 363
نقد النماذج ---- 368

اسلوب نظرية القرار في الاختيار ———— 369
الخلاصة ———— 375
التمارين ———— 377

الفصل الثالث عشر: التحليل العاملي ———— 382
مثال بياناته افتراضية ———— 383
العوامل وتشبعاتها ———— 383
الدوران ———— 385
العوامل المرتبطة ———— 389
عدد العوامل ———— 390
نموذج التحليل العاملي ———— 392
الطائفية والفردية ———— 393
مثال بياناته حقيقية ———— 394
التحليل العاملي الاستكشافي والتأكيدي ———— 404
الخلاصة ———— 406
التمارين ———— 407

4 الوحدة الرابعة: تحليل الفقرات في تطوير الاختبار

الفصل الرابع عشر: تحليل الفقرات ———— 413
صعوبة الفقرة والمتوسط والتباين ———— 413
تمييز الفقرة ———— 416
معاملات ثبات الفقرة وصدقها ———— 423
اجراء دراسة تحليل الفقرات ———— 425
تحليل الفقرات في الاختبارات محكية المرجع ———— 434
الخلاصة ———— 441
التمارين ———— 443

الفصل الخامس عشر: مدخل الى نظرية الاستجابة للفقرة 450
مفاهيم أساسية لنظرية الاستجابة للفقرة 451
المنحنى الطبيعي 459
ربط نظرية الاستجابة للفقرة بنظرية القياس التقليدية 463
النماذج اللوغاريتمية 466
حساب المعالم 469
اختيار النموذج 477
تطبيقات نظرية الاستجابة للفقرة 479
الخلاصة 490
التمارين 491

الفصل السادس عشر: اكتشاف تحيز الفقرة 498
طرائق تعتمد نظرية الاستجابة للفقرة 499
تقنيات مربع كاي 507
مزايا الطرائق المختلفة ومساوئها 518
الخلاصة 519
التمارين 520

5 الوحدة الخامسة: درجات الاختبار وتفسيرها

الفصل السابع عشر: التصحيح من اثر التخمين وطرائق التصحيح الاخرى 525
صيغة التصحيح 526
مكافأة المعرفة الجزئية 532
الخلاصة 536
التمارين 537

الفصل الثامن عشر: اعداد المعايير 540
اساليب إعداد المعايير 541
البحوث التجريبية في اساليب اعداد المعايير 548

الاعتبارات التطبيقية في إعداد المعايير ـــــــــــــــــــــــــــــــــــــــ 551
الاعتبارات التقنية في اعداد المعايير ـــــــــــــــــــــــــــــــــــــــ 553
الخلاصة ـــــــــــــــــــــــــــــــــــــــ 563
التمارين ـــــــــــــــــــــــــــــــــــــــ 564

الفصل التاسع عشر: المعايير والدرجات المعيارية ـــــــــــــــــــــــــــــــــــــــ 568
اجراء دراسة اعداد المعايير ـــــــــــــــــــــــــــــــــــــــ 569
المعاينة الاحتمالية ـــــــــــــــــــــــــــــــــــــــ 569
وصف دراسة المعايرة في دليل الاختبار ـــــــــــــــــــــــــــــــــــــــ 576
انواع الدرجات المعيارية ـــــــــــــــــــــــــــــــــــــــ 578
الخلاصة ـــــــــــــــــــــــــــــــــــــــ 594
التمارين ـــــــــــــــــــــــــــــــــــــــ 595

الفصل العشرون: معادلة درجات الاختبارات المختلفة ـــــــــــــــــــــــــــــــــــــــ 602
جمع البيانات للمعادلة ـــــــــــــــــــــــــــــــــــــــ 603
المعادلة بوساطة نظرية الاستجابة للفقرة ـــــــــــــــــــــــــــــــــــــــ 615
معادلة الدرجات الحقيقية ـــــــــــــــــــــــــــــــــــــــ 622
المكافأة في المعادلة ـــــــــــــــــــــــــــــــــــــــ 629
الخلاصة ـــــــــــــــــــــــــــــــــــــــ 630
التمارين ـــــــــــــــــــــــــــــــــــــــ 631

المراجع ـــــــــــــــــــــــــــــــــــــــ 635
الملحق أ : الاحتمالات وقيم الاحداثي ص المرتبطة بدرجات زائية معينة تحت المنحنى
المعياري الطبيعي ـــــــــــــــــــــــــــــــــــــــ 653
الملحق ب: دليل حل التمارين ـــــــــــــــــــــــــــــــــــــــ 655
الملحق جـ: مفاهيم ومصطلحات ـــــــــــــــــــــــــــــــــــــــ 681

الوحدة الأولى

مدخل الى نظرية القياس

الفصل الاول

1

ما هي نظرية القياس

الفصل الاول

ما هي نظرية القياس

أي شيء موجود فإنه موجود بمقدار، ولمعرفته بصورة شاملة أعرف كميته بالإضافة إلى نوعيته (Thorandike, 1918).

يواجه الطالب المبتدىء في دراسة نظرية القياس بأسئلة من مثل:

- ما هي نظرية القياس ولماذا نحتاجها؟
- ما الإختلاف بين تعلم تطبيق اختبار عن استخدام اختبارات بعينها؟
- هل أن أصول نظرية القياس في علم النفس أم أنها مجال دراسة جديد؟
- كيف ترتبط نظرية القياس بتصميم البحث وإحصائياته؟

يجيب هذا الفصل جزئياً على الأقل عن مثل هذه الأسئلة، وفي الأجزاء التالية سنحدد المفاهيم الأساسية، وسنصف مشكلات معينة تواجهنا في قياس الأبنية النفسية، وتفسير الهدف الواسع لنظرية القياس بالنسبة لمشكلات مثل هذه، ومراجعة الأصول التاريخية لهذا الميدان، وأخيراً سنوضح الفروق بين نظرية القياس والمظاهر الأخرى لأبحاث العلوم الاجتماعية وتقويمها.

لفهم الحاجة إلى نظرية القياس، من الضروري أولاً فهم بعض الأمور حول الطبيعة الأساسية للقياس، والأبنية والاختبارات النفسية. وقد وصف ويتزنوفر (Weitznhooffer, 1951) القياس بأنه عملية يجريها المجرب على العالم الطبيعي. وعرّف ستيفنز (Stevens, 1946) القياس على أنه إعطاء أو تأشير بالأرقام للأشياء والأحداث وفق قواعد معينة، ودعم كل من لورد ونوفيك (Lord & Novick, 1968) وتورغرسون (Torgerson, 1958) دقة تعريف ستيفنز وأضافوا بأن القياس يكون لخصائص الأشياء لا الأشياء نفسها. فعندما يقيس عالم الكيمياء الفيزيائية الكتلة الجزيئية لمركب أو يحدد عالم الحياة عدد البكتيريا في نقطة ماء، فإن هذه تعد قياسات لخصائص معينة للأشياء التي تم قياسها. وبصورة مماثلة، فإن عالم النفس لا يقيس الطفل ولكنه يقيس صفات طبيعية معينة مثل الطول أو الوزن أو خصيصة نفسية مثل تطور الألفاظ، أو النضج الإجتماعي أو عملية الجمع الأساسية. وعُرِّف علم النفس -تقليدياً- على أنه دراسة السلوك والخصائص التي تُميز الفرد في المنزل والعمل والمدرسة والمواقف الاجتماعية، وهذه

يطلق عليها اسم الخصائص النفسية (أو السمات النفسية).وعلى عكس الخصائص الفيزيائية، فإن الخصائص النفسية للفرد لا يمكن قياسها مباشرة كما هو الحال في قياس الوزن أو الطول. فالخصائص النفسية عبارة عن أبنية، إنها مفاهيم افتراضية -نتاج لما كوّنه الخيال العلمي لعلماء الاجتماع الذين حاولوا تطوير نظريات تفسر سلوك الإنسان. ولا يمكن إثبات وجود مثل هذه الأبنية على الإطلاق، وعلى العكس فإن الدرجة التي تميز الخصيصة النفسية عند فرد ما يمكن استنتاجها -فقط- من الملاحظات السلوكية لهذا الفرد.

دعنا نطرح المثال التالي: عملية تكوين البناء وكيف يؤدي إلى قياس الخصائص. افترض أن عالم نفس تطوري يعمل مع أطفال ما قبل المدرسة ولاحظ من خلال أنشطة اللعب المتعددة بأن بعض الأطفال يحاولون وبصورة متكررة توجيه نشاط الآخرين. وبعد ملاحظة هذا النوع من السلوك المتسق عبر الزمن وخلال سياقات عديدة لمجموعة الأطفال نفسها، فإن عالم النفس يبدأ بتسمية مثل هذه السلوكات على أنها «السيطرة الإجتماعية» وعالم النفس هذا ابتكر بناءاً نظرياً يشمل عدداً من السلوكات المتشابهة، ولكنّ ابتكار البناء لا يكون مشابهاً لقياس البناء نفسه. فقبل إجراء أي قياس للبناء؛ من الضروري تكوين بعض قوانين التطابق بين البناء النظري والسلوكات الملاحظة التي تعد مؤشرات منطقية لهذا البناء، وتسمى هذه العملية تكوين أو تأسيس التعريف الإجرائي للبناء. وفي مثالنا لقياس السيطرة الاجتماعية؛ يجب أن يحدد عالم النفس أنواع سلوكات أطفال ما قبل المدرسة التي يمكن بناءاً عليها اعتباره «مسيطراً» وبعدها عليه أن يقترح خطة للحصول على مثل هذه السلوكات في مواقف معيارية، وتسجيل ملاحظات عن كل طفل بصيغة معيارية. وهذا يتطلب تطوير أداة أو أختبار للبناء الذي يعرف بالسيطرة الاجتماعية. وبشكل عام يمكن تعريف الاختبار على أنه طريقة معيارية للحصول على عينة من السلوك في مجال معين. كذلك فإن هذا الكتاب سيستخدم مصطلح الاختبار ليشير إلى الطرائق المتبعة للحصول على عينة أقصى أداء للفرد (كما هو الحال في اختبارات الإستعداد والتحصيل حيث يطلب من المفحوص تقديم أفضل ما لديه) أو الأداء المميز للفرد (كما هو الحال في الاستبانات أو المقابلات حيث يكتب المستجيب مشاعره وإتجاهاته وميوله وتفاعله مع المواقف المميزة). وأسلوب آخر في الحصول على عينة أداء مميز هو قائمة معيارية وقائمة بالسلوكات التي يمكن أن يستخدمها الملاحظ ويسجل فيها سلوكات الفرد في مواقف طبيعية. وعندما يُعد عالم النفس قائمة بسلوكات ليتم فحصها أو تقديرها من قبل ملاحظين حسب خطة محددة مسبقاً، وعندما يضع الكيميائي مجموعة فقرات اختيار من متعدد لمستوى نشاط العناصر في الجدول الدّوري، وعندما يُقّدر المرشد النفسي درجة القلق التي يشعر بها المسترشد في المواقف الاجتماعية، فكل واحد من هؤلاء يقترح اختباراً في

المجال الواسع لهذه الكلمة، ونحن سنستخدم هذا المصطلح بالطريقة نفسها.

ويظهر قياس الخصيصة النفسية عندما نؤشر قيمة كمية لعينة سلوكية جُمعت بواسطة اختبار. وبعبارة أخرى، يحدث القياس عندما يخبر عالم النفس بعدد مرات السيطرة الإجتماعية من خلال قائمة توضيح عدد مرات ظهور السلوك خلال فترة خمس دقائق أو عندما يعد الكيميائي عدد الفقرات التي أجاب عليها الطالب إجابة صحيحة ويسجل درجة كلية. ومن قياسات كهذه لسلوكات ملاحظة يكوِّن مطور الإختبار إستدلالاً حول كمية البناء النظري والتي تميز فرداً بعينه.

لماذا يحاول التربويون والنفسيون قياس أبنية لايمكن ملاحظتها مباشرة؟ في التربية وعلم النفس، يزودنا البناء بطريقة فعّالة وملائمة لتصنيف عدد من السلوكات المتشابهة. وبدون مثل هذه الأبنية التي تستخدم في وصف سلوكات الأفراد الذرية وتصنيفها، فإن معظم المحاولات لملاحظة الظواهر السلوكية المعقَّدة ستذوب في التشويش والغموض، خذ مثلاً المدى الواسع من الكلمات والأفعال والتفاعلات التي يمكن أن تظهر في يوم واحد في صف دراسي واحد. فقد يكون كرونباخ (Cronbach, 1969) قدّر هذه الحالة عندما أكد على أن دراسة الأنشطة الصفية هي على الأقل بصعوبة دراسة الإعصار. والملاحظ الذي يحاول تسجيل ووصف أفعال كل فرد في هذا الموقف سيرتبك فوراً، علاوة على أن أي محاولة للتنبؤ أو ضبط أو تحديد مسبب كل سلوك منفصلاً عن غيره سيكون بلا أمل أيضاً. ويمكن للملاحظ من خلال استخدامه للأبنية أن يبدأ بتصنيف وتجميع السلوكات المتشابهة ثم وصفها بمصطلحات مطابقة لخصائصها. وتعد مثل هذه الأبنية اللبنات الأساسية في بناء النظريات حول السلوك الإنساني. وعلى المستوى الأبسط؛ فإن النظرية النفسية هي صياغة للعلاقات الممكنة بين بناءين نفسيين أو بين البناء والظواهر الملاحظة من النتائج التطبيقية. وتحديداً، من خلال الفحص التجريبي والتجريد باستخدام النظريات يصبح من الممكن التنبؤ بنماذج سلوكية معينة أو ضبطها. ولتحقيق هذا الهدف، فمن الضروري تكميم الملاحظات السلوكية التي تمثل الأبنية التي وضعها المنظرون.

مشكلات في قياس الأبنية النفسية:

لأن الأبنية النفسية مجردة وتقيَّم -فقط- بصورة غير مباشرة، فإن تصميم أدوات لقياسها يتضمن مشكلات عديدة. افترض على سبيل المثال؛ افترض الرغبة في قياس الأبنية الخاصة بالمواقف الثلاثة الموضحة أدناه:

1. عالم نفس شخصية يريد تطوير اختبار في الاستعداد الميكانيكي لاختيار موظفين تقدموا لوظيفة في موقع صناعي.
2. عالم نفس مدرسي يريد تطوير مقياس لتقييم اتجاهات المعلمين نحو الطلبة المعوقين حركياً.
3. معلم الصف الخامس يريد تطوير اختبار لقياس مهارات القسمة الطويلة.

ومع أن الأبنية الخاصة بكل من الإستعداد الميكانيكي والاتجاهات نحو المعوقين ومهارات القسمة الطويلة تعد متباينة، إلا أن مُطور الإختبار سيواجه الخمس مشكلات القياسية على الأقل، وهي شائعة في القياسات النفسية كافة.

1. **لا توجد طريقة واحدة لقياس أي بناء موافق عليها عالمياً**، وذلك لأن هذه القياسات غير مباشرة وتعتمد على سلوكات يعتقد أنها مناسبة للبناء قد الدراسة، فمن الممكن أن يتحدث منظرين أثنين عن البناء نفسه باستخدام أنواع مختلفة من السلوكات في تحديد البناء إجرائياً. دعنا نأخذ على سبيل المثال تطوير إختبار قياس مهارة الطلاب في القسمة الطويلة. وحيث أنه لا يمكن لأحدنا أن ينظر إلى رأس الطالب ويرى كم يعرف عن القسمة الطويلة، فإن مُطور الاختبار يصمم بعض السلوكات التي يمكن أن يجريها الطلبة ليستدل منها عن مدى معرفتهم للموضوع. وإحدى الطرق هو تكليف الطالب بحل عدد من المسائل المتعلقة بالقسمة الطويلة، وطريقة ثانية هي سؤالهم عن وصف الخطوات المتسلسلة المستخدمة في عملية القسمة الطويلة، وطريقة ثالثة هي سؤالهم عن اكتشاف الأخطاء في مسألة قسمة طويلة. من الواضح إذن أن عمليات قياس مختلفة تنتج عن تعريفات إجرائية مختلفة تؤدي بالطبع إلى إستنتاجات مختلفة عن مستوى معرفة الطلبة.
2. **تعتمد القياسات النفسية على عينات محددة من السلوك.** ففي المثال السابق، من المستحيل تعريض الطلبة للمسائل الممكنة جميعها في القسمة الطويلة والتي يتوقع أن يحلها الطالب. لذلك. فإن أي محاولة لقياس المهارات بهذا المجال يجب أن تستخدم عينة من هذه المسائل جميعها. ولتوفير عينة مناسبة من المجال السلوكي، فإن تحديد عدد الفقرات وتباين المحتوى يعد أمراً ضرورياً.
3. **القياس الناتج مُعرَّض للخطأ دائماً**، وذلك لأن معظم القياسات النفسية تعتمد بالأساس على عينة محدودة من الملاحظات وتؤخذ عادة في وقت واحد فقط. دعنا نعود لمثال القسمة الطويلة السابق، فإذا تقدم الطلبة للإختبار نفسه مرتين متتاليتين، فمن غير المعتاد أن يحصلوا على الدرجات نفسها بسبب تأثير الموقف، والملل، والنسيان،

والتخمين، وإهمال الإشارات أو نسيانها. وفي حالة تقدمهم لصيغتين اختباريتَيْن مختلفتين؛ فمن المحتمل أن لا تتشابه درجاتهم بسبب الإختلاف في المحتوى إضافة إلى العوامل السالفة الذكر. ومثل عدم التطابق في الدرجات قد يكون بسبب عينة المهام أو الأحداث المسببة للخطأ. لذلك؛ فإن المشكلة القائمة في القياس النفسي تكمن في كيفية حساب درجة الخطأ الموجودة في مجموعة الملاحظات أو القياسات المتوافرة.

4. **عدم وجود وحدات محددة بالضبط على المقياس يؤدي إلى مشكلة أخرى.** فهل حقيقة أن المفحوص الذي لا يستطيع الإجابة على أي فقرة إختبارية في القسمة الطويلة يعني أن درجة سيطرته هلى هذه المهارة = صفر؟ وإن أجاب سو على (5) فقرات وجو على (10) فقرات وستيف على (15) فقرة إجابة صحيحة هل يمكننا الإفتراض أن فرق الكفاية بين سو وجو هو الفرق نفسه بين جو وستيف؟ وهل يكون انتشار الطلبة الثلاثة متساوٍ على متصل القدرة التي يقيسها الإختبار؟ إن تحديد خصائص التدريج، واشتقاق وحدات القياس، وتفسير القيم المشتقة منه تعد قضايا معقدة يجب أن تؤخذ بعين الاعتبار عند تطوير أية أداة نفسية واقتراح نظام للتّصحيح وإعطاء الدرجات.

5. **لا يجب تحديد الأبنية النفسية بصيغة إجرائية فقط،** ولكن يجب تبيان علاقاتها مع الأبنية الأخرى أو الظواهر الملاحظة. ومع أن القياس النفسي يعتمد على الاستجابات الظاهرة أو الملاحظة فإن معناه يكون قليلاً أو عديم الفائدة حتى يتم تفسيره في ضوء البناء النظري الذي يقع ضمنه. ولهذا السبب شدَّد كل من لورد ونوفيك (1968) على أهمية تحديد الأبنية المتضمنة في القياس النفسي على مستويين اثنين هما:

الأول: كما لاحظنا، فإنه يمكن تعريف البناء من خلال السلوك الملاحظ، وهذا النوع من التعريف يحدد الكيفية التي تجري بوساطتها عملية القياس.

الثاني: يمكن تعريف البناء من خلال العلاقة المنطقية أو الرياضية مع الأبنية الأخرى ضمن الإطار النظري. وهذا النوع من التعريفات يزودنا بأساس في نفسير القياس الناتج. وإن كان من غير الممكن إثبات مثل هذه العلاقة تجريبياً فإن القياس الناتج لا قيمة له. فالحصول على أدلة تبين كيفية ارتباط قياس نفسي بقياس أبنية أخرى أو أحداث في العالم الواقعي، يبقى التحدي الأساسي في تطوير القياس.

نظرية القياس على أنها فرع من فروع المعرفة.

تعرف دراسة رمشكلات القياس السابق ذكرها وحلولها التي ظهرت في العلوم التربوية والنفسية باسم نظرية القياس. ويبحث محتوى مساق في نظرية القياس أساساً في:

1- تقدير مدى تأثير هذه المشكلات على قياس أُجري في موقف معين.

2- اقتراح وسائل أو طرائق للتغلب على هذه المشكلات أو تقليل أثرها.

إن هدف مساق نظرية القياس هو مساعدة الطلبة لإدراك النماذج المنطقية والرياضية التي تشكل الأساس في بناء الإختبارات المعيارية وتطبيقاتها. ويؤدي إدراك هذه النماذج وافتراضاتها إلى تطبيقات مطوّرة في اتخاذ القرارات. ومن المهم تمييز نظرية القياس عن موضوع التقييم التربوي والنفسي الأكثر شيوعاً وتطبيقاً، والذي يركز عادةً على إدارة اختبارات معينة وتفسيرها. بالمقابل؛ فإن نظرية القياس تزودنا بهيكل عام يصوِّر عملية تطوير الأداة. فالنماذج النظرية والطرائق التقليدية التي تُدَرَّس في مساق التي تُدَرَّس في مساق مثل هذا لا تقف عند نظرية نفسية أو تربوية معينة، بل تكون جميعها مفيدة وبدرجات متساوية في قياس العديد من الخصائص النفسية.

وللعديد من التصميمات التجريبية والتحليلات الإحصائية المستخدمة في الأبحاث التربوية والنفسية لها أصولها في الأبحاث الزراعية وذلك عندما يكون من الضروري تقييم أثر معالجة تجريبية في الظروف الواقعية لا في المختبر. وقد تم اشتقاق مبادئ نظرية القياس لتلبي حاجات القياس المتعلقة بمجالات العلوم التربوية والاجتماعية.

الأصول التاريخية:

يتماشى التطور التاريخي لنظرية القياس مع تطور علم النفس كفرع من فروع المعرفة العلمية. وظهرت النظرية من خلال جهور علماء النفس في أوروبا والولايات المتحدة إذ درسوا المشكلات التّربوية والنفسية المتنوعة. وفي هذا الجزء سنذكر باختصار مساهمة بعض الأفراد الذين ساهمت جهودهم في تقديم نظرية القياس وتطبيقاتها.

المانيا: والبدء في منتصف عام 1800م، وفي مختبرات الإدراك في ليبزغ إذ كان ويليم فوندت وإرنست ويبر وغوستاف فخنر من أوائل الذين ثمنوا أهمية الحصول على قياس نفسي ضمن طروف مضبوطة بدقة. فقد كانت دراسة علم النفس في السابق نوعاً من الاستبطان الفلسفي والملاحظات غير المنهجية. وكتب فوندت - مؤخراً- (Wundt, 1873) يصف الثورة على الطريقة السابقة يقول «في الوقت الذي يرى الباحث إستخدام منهجية دقيقة في الإجابة على أي سؤال في علم النفس تتحداه الفلسفة عند كل نقطة على أنه علم طبيعي لإثبات شرعية المحاولة». ومع أن قياس القدرات العقلية توسع من خلال قياس زمن الرجع والتمييز الحسي وتقدير الأوزان النسبية للأشياء أو غيرها من المتغيرات الحسية التي تمت دراستها في مختبرات الإدراك الألمانية. وتبدو شرعية جميع الملاحظات السلوكية تحت شروط محدَّدة -

مسبقاً– واضحة من خلال التعليمات الدقيقة التي ترافق الإختبارات المقننة الحديثة، فمثل هذه التعليمات تحدد الشروط المعيارية لجمع البيانات ونظام التصحيح وجعل هذه الدرجات تتطابق مع التعريف العام للقياس.

بريطانيا: خلال الفترة نفسها انشغل العلماء البريطانيون في العمل الذي له أثر ملموس في قياس السمات العقلية. وبعكس الألمان انصبّ اهتمامهم على دراسة الفروق الفردية، ومن أشهر هؤلاء فرنسيس جالتون الذي تأثرت أفكاره تأثراً واضحاً بتفكير خاله شارلس دارون.

وتبدو إهتمامات جالتون وأساليبه مختلفة بعض الشيء وكأنها لا أرثوذكسية بمعايير اليوم. فمثلاً وصف في إحدى مقالاته بعض التطبيقات غير الرسمية مثل عدد الأشياء التي يمكن أن يتذكرها عندما كان يتجول في البول مول (مركز تجاري كبير) أو بعد ملاحظته قطيع حيوانات يجفل أو يُروَّع عند سماعه صفارة عالية مخفية في مقبض عصا (Galton, 1883). وقد سجل أيضاً قياسات كمية مفصلة للخصائص النفسية للأفراد في مختبر الأنثروبولوجيا وبعد إستخدام التطبيقات والتقنيات الإحصائية للبيانات المستمدة من الإختبارات النفسية مساهمته الرئيسية في مجال نظرية القياس. وقد استخدم في عام 1869 درجات اختبار الرياضيات لطلبة جامعة كمبردج جميعهم ودرجات امتحان القبول للكلية العسكرية الملكية في إثباته أن القدرات العقلية تتوزع اعتدالياً على وجه التقريب. وبعد ذلك العشرين عاماً ظل مبهوراً بطبيعة الفروق الفردية وقد كتب مقالة اقترح فيها استخدام الطرائق الإرتباطية في اختبار التباين المشترك بين سمتين أو أكثر (Dennis, 1948). وقد طور الإحصائي البريطاني كارل بيرسون صيغة إحصائية لمعامل الإرتباط تعتمد على هذا الإقتراح. وتبعه علماء آخرون ساروا على النهج نفسه. والأكثر شهرة بين هؤلاء هو شارلس سبيرمان بنظريته المعروفة عن الذكاء والتي تعتمد على طريقة إرتباطية أكثر تطوراً تسمى التحليل العاملي. وبقيت الطرائق الإرتباطية والتحليل العاملي التقنيات الأكثر شيوعاً واستخداماً في إجراءات تصديق الإختبارات.

فرنسا: على الرغم من إسهامات النّفسيّين الألمانيّين والبريطانيّين المهمة السابقة، إلا أن إنجاز الفرنسيين الفرد بينيه وثيوفيل سيمون (1905-1908) اللّذيْن نقلا دراسة الإختبار العقلي من تدريب أكاديمي إلى مغامرة أصبح لها تطبيقات فورية في غرفة الصف، وفي العيادات النفسية، وفي أماكن العمل. فقد أخذ هذان المحلّلان النفسيّان على عاتقهما المهمة العملية في إبتكار طريقة تفيد في تحديد الضعف العقلي عند أطفال المدارس العامة. وكانت النتيجة رائعة في قياس السمة التي تعرف بشكل عام باسم «الذكاء» ومع أن شهرة بينية كانت بسبب اختبار الذكاء الذي ما زال يحمل أسمه، إلا أنه أثّر تأثيراً مهماً في نظرية القياس وذلك في

فرنسا: بينيه وسيمون: صمم الاختبار العقلي تطبيقات تم في المدارس بدلاً من تدريب أساتذة
٢- طريقة في تحديد الضعف العقلي لدى تلاميذ المدارس
٣- قياس مستوى الذكاء
٤- طريقة عملية بناء الاختبار وصدقه: اعطاء العديد من المهام والفقرات للأطفال
المهام التي ميزت بين الأطفال تم تجميعها في مستويات عمرية
٥- تطوير مفهوم المعايير
٦- تقدير العمر العقلي بالمقارنة مع العمر الزمني
الصيغة الحالية لاختبار ستانفورد بينيه تشبه الصورة الأصلية

إنشائه لطريقة عملية في بناء الاختبار وتصديقه. وحتى إختراق بينية للطريقة الشائعة المتبعة في بناء الإختبار والتي هي تجميع فقرات بناءاً على منطق نظري ومن ثم إستخدامها، ففي عام (1905) رفض بينية هذا الأسلوب وسماه بالعمل المسلي (المضحك) مقارنة بالإستيطان وبناء المستعمرة في الجزائر بالإعتماد على الخريطة فقط ودون معرفة العنوان. فطريقة بينية الجادة تتطلب تحديد سلسلة من المهام التي تميز بين الأطفال من الفئات العمرية المختلفة وعبر التطور المتسلسل، وقد تم تطبيق مئات الفقرات على اعداد كبيرة من الأطفال اختير من بينها الفقرات التي مَيَّزت جيداً بين الأطفال، وبعد ذلك قام بينية وسيمون بتجميع مهام متنوعة لكل مستوى عمري، ويؤديها معظم أطفال ذلك المستوى بنجاح، فمثلاً:

في عمر 3 سنوات: يؤشر إلى العين، الأنف، الفم.

في عمر 7 سنوات: يشير إلى المحذوف في صورة، يكرر خمسة أشكال رآها.

وفي عمر 11 سنة: ينتقد الجمل التي تحوي تناقضات.

ومن الأداء على الفقرات اشتق بينية تقديراً للعمر العقلي للمفحوص مقارنة بالعمر الزمني له، وذلك لتحديد المكان المناسب له في المواقف التدريسية أو في المؤسسات الخاصة بالمعاقين. وإضافة إلى تأسيس بينية لعملية التحليل الأولي للفقرات فإننا ندين له في تطويره لمفهوم المعايير التي تعد مرشداً مهماً في تفسير الدرجات ويعد تقديراً لجهود بينية وبراعته أن الصيغة الحالية لاختبار ستانفورد – بينيه للذكاء لا تزال تشبه بصورة ملحوظة الصورة الأصلية للإختبار التي بنيت أصلاً لأطفال المدارس في باريس (Thorndike, 1975). ومن المهم ملاحظة أن تحليل الفقرات وإعداد المعايير، وعلى الرغم من التقنيات؛ فإنها تشبه كثيراً طريقة بينية التي أسسها في مطلع هذا القرن.

الولايات المتحدة: مع أن اهتمام التربويّين الأمريكان للمشكلات النفسية كان منذ مطلع 1800م، إلا أنهم لم يبدأوا بتوفير طرق قياس متميزة حتى بداية القرن العشرين. فمثلاً يعود الفضل إلى جيمس كاتل بابتكار مصطلح الإختبار العقلي وذلك في مقالة كتبها عام 1890م. فقد شدد في كتاباته في تلك الفترة على أهمية اختيار عينات اختبارية كبيرة وذلك للحصول على تقديرات طبيعية دقيقة، ويصر كذلك على أن علم النفس يجب أن يدرس طبيعة أخطاء الملاحظة (Boynton, 1933)، مع العلم بأن الاختبارات التي وصفها كاتل هي نفس حركية أو إدراكية وتشبه كثيراً تلك المستخدمة من قبل علماء النفس التجريبي الألمان. وألف ثورندايك عام 1904 أول كتاب في نظرية القياس: «مقدمة في نظرية القياس العقلي والاجتماعي»، وكان غير متأكدٍ من ردود أفعال زملائه، وقدم نسخة منه إلى زميله وليم جيمس مع الملاحظة الآتية: (Joncich, 1968, p. 290):

«أرسل أليك كتاباً يثير الرهبة قمت بتأليفه، وبدون كلام فهو علمي، وخالٍ من أي اهتمام

إنساني يمكن من خلاله إجراء أبحاثك جميعها على الإنسان، ولكن لا تنظر إلى أنه يعنيك أنت نفسك، فالأشكال والمنحنيات والصيغ قد تقود إلى الجنون أو الحمق». ومع ذلك؛ فإن محاولة ثورندايك بأن يضع من الآن فصاعداً تفكيراً منظماً للمشكلات القياسية في علم نفس نودي بها على صعيد واسع، وأصبحت دراسة نظرية القياس متطلباً لتدريب العديد من النفسيين الشباب.

ومن قاعات المحاضرات في كولومبيا وهارفارد وستانفورد، استقطب الباحثون من أمثال ثورندايك وكاتل تخيلات أو تصورات الزملاء الشباب والطلبة الذين يمثلون الجيل التالي من النفسيين سواء المنظرين منهم أو التجريبيين، فقد كيّف مطوروا الاختبارات عمل بينية وباهتمام بالغ في إنتاج ضخم يلائم المجتمع الأمريكي خلال هذه الفترة من الثورة الصناعية، وأنتجوا أول مجموعة من اختبارات الذكاء. ففي عام 1917 وبدخول أمريكا الحرب العالمية الأولى؛ قامت لجنة من علماء النفس (بنجهام، وغودارد، وهاينز، وتيرمان، وويل، وواييل وييركس) بإنتاج خمسة صيغ لاختبارات يتقدم إليها المستخدمين في العسكرية جميعهم، وقد منح هؤلاء راتباً أساسياً يبلغ 700 دولار من قبل قسم الحرب، وبعدها أعدّ هؤلاء مجموعة اختبارات غير لفظية للمفحوصين الذين لا يتحدثون الإنجليزية. واستخدمت هذه الاختبارات في اختيار وتعيين الأفراد الجدد، وفي تحديد المرشحين للتدريب، وفي تسريح الأشخاص غير الكفوئين عقلياً (Yerkes, 1921). وعلى الرغم من النجاح في الإستخدامات العسكرية؛ إلا أن علماء النفس كرسوا جهودهم لاستخدام اختبارات الذكاء مع المدنيين في الإختيار التربوي والمهني. ويجب ملاحظة أن جهودهم أظهرت صراعاً بين مؤيد ومعارض لاستخدام الإختبارات المعيارية والتي استمرت أكثر من خمسين عاماً (Cronbach, 1975).

وأضاف كل من ثيرستون وشيف (Thurstone & Chave, 1929) تقنيات لتدريج مقاييس الإتجاهات إلى الأدب المتنامي في تطوير الإختبارات. وأضاف كل من ثيرستون وكيلي وهولتزنجر طريقة جديدة في التحليل العاملي. وفي الثلاثينات؛ حدثت تطورات متقدمة بسرعة هائلة لتدعم صناعة الإختبارات النامية إذ أن العلماء في هذا المجال أسسوا مجتمعاً نفسياً يُسهِّل الإتصال بين الباحثين من خلال مجلة أطلق عليها أسم سيكومتريكا (Psychometrika). وتبعها في عام (1943) نشر مجلة تطبيقية بدرجة أكبر هي القياس التربوي والنفسي (Psychological and Educational Measurement). ومع نشر مثل هذه المجلات المتخصصة والعديد من المؤلفات في هذا الميدان؛ فإن نظرية القياس يبدو أنها أصبحت أحد فروع المعرفة التي تهم التربويّين والنفسيّين.

دور نظرية القياس في البحث والتقويم:

لتوضيح دور نظرية القياس ضمن الشبكة الأوسع في منهجية البحث والتقويم، فمن الضروري اعتبار الأبحاث في العلوم التربوية والإجتماعية على أنها عملية سؤال أو اختبار

يتألف من مراحل بصورة جيدة، وهذه يمكن تمييزها على أنها:

1. صياغة سؤال أو فرضية البحث.
2. تحديد التعريفات الإجرائية لكل متغير في الفرضية، وذلك بتحديد كيف يمكن ضبطه أو قياسه خلال البحث.
3. تطوير أو اختيار أداة وطريقة لتكميم الملاحظات على المتغيرات جميعها.
4. اختبار دقة وحساسية الأدوات والطرق المستخدمة.
5. تجميع البيانات التجريبية ضمن هيكل التصميم التجريبي، وهذه تمكننا من الإجابة على السؤال الأصلي.
6. تلخيص البيانات رياضياً، وعندما تكون ملائمة إجراء الإختبارات الإحصائية لتحديد أرجحية ما إذا كانت البيانات الملاحظة تعود للصدفة.

إن قناعات نظرية القياس تتلائم بدرجة عالية مع الخطوات (من 2 إلى 4) في هذه العملية. ومن الأساسي ملاحظة أن الخطوة (4) تتطلب تجميع البيانات وتلخيصها إحصائياً من خلال عينة استطلاعية، وذلك للتأكد من أن الأداة تعمل بفعالية قبل إجراء التجريب أو إجراء دراسة تكوين التدريج (الخطوات 6,5). وتعزى النتائج الغامضة والمتعارضة التي تُميِّز الأبحاث في العلوم التربوية والإجتماعية إلى الفشل في التثبت من أداة البحث أثناء تطويرها واختبارها – مسبقاً– قبل إجراء البحث.

ففي العلوم الطبيعية والحياتية، تختبر طرائق القياس قبل استخدامها في أي تجربة. ولا يحاول الكيميائي الحيوي تحليل أي مركب مجهول في المختبر دون استغراق الوقت الكثير في معايرة السبكترومتر للتأكد من أنه يعطي قراءات دقيقة لطول الموجة للعينة المعيارية. بالمقابل؛ فإن علماء الاجتماع يحاولون تقييم كيفية تأثير برنامج تجريبي على متغير معقد مثل مستوى القلق العقلي للمريض أو تطور لغة طفل ما قبل المدرسة باستخدام اختبارات لم تُجرَّب عملياً.

إن الباحث الذي يقفز مباشرة من صياغة سؤال البحث وتحديد المتغيرات إلى جمع البيانات التجريبية يشبه من يذهب بأيدٍ فارغة. فهل يلاحظ الباحثون أنه لا فرق بين المعالجات التي لا يستطيعون من خلالها معرفة أن المعالجات كانت غير فعّالة أو أن القياسات كانت غير دقيقة بحيث أنه لم يتم اكتشاف الأثر الحقيقي للمعالجة. إن التأكيد والمناداة بمبدأ بناء الإختبار، وتجريب الأداة قبل إجراء التجربة النهائية سيؤدي حتماً إلى تطويرٍ نوعيّ شاملٍ للأبحاث التربويّة والنفسيّة.

تنظيم هذا الكتاب:

إن معرفة نظرية القياس ضرورية لأي باحث أو مقوِّم تتطلب تجاربه تطوير اختبار نفسي لقياس المتغير الذي يهتم بدراسته. ولهؤلاء الذين يستخدمون الإختبارات الصفية وغيرهم من المحللين النفسيين، فإن بعض المعرفة لنظرية القياس ضرورية لهم. إن اختيار أداة مُطوَّرة جيداً والتفسيرات المنافسة للدرجات هي مسؤولية مستخدمي الاختبارات جميعهم. فالفهم للنماذج النظرية المعتمدة في بناء الاختبار تُمكِّن المطبق من تحقيق هذه الواجبات المهنية المتخصصة. وهدف هذا الكتاب هو تأسيس تطبيقات مناسبة لاستخدام الإختبارات وتطويرها.

وقد تم تنظيم مادة الكتاب ضمن خمس وحدات مميزة ومرتبطة -معاً- بالوقت نفسه. فالوحدة الأولى تركز على المفاهيم المدخلية للقياس والتدريج والتي يجب أن تكون مألوفة لكل من يسخدم اختباراً أو يطوِّره. والوحدتين الثانية والثالثة تتعلق بالخصائص المهمة لدرجات الإختبار (الثبات والصدق). والتي تعد أساسية فيما لو استخدمت درجات الإختبار في اتخاذ قرارات سواء على مستوى الفرد أو المجموعة. وتم عرض النظرية السيكومترية التي تشكل الأساس والطرائق المستخدمة في فحص ثبات الدرجات وصدقه بعمق. ومن الواضح أن معرفة كهذه تعد ضرورية لُطَوِّر الإختبار، وكذلك من الضروري لمستخدمي الاختبارات أن يكونوا على إطلاع جيد بهذه الموضوعات، لذا يمكنهم التقييم الناقد للإختبارات المنشورة، وما إذا كان تطويرها وتصديقها ضمن معايير متخصصة ومقبولة. والوحدة الرابعة مكرّسة لطرائق خاصة للفحوص التجريبية لنوعية الفقرات التي تعد مهمة في تطوير اختبارات جديدة أو مراجعة وتحسين الإختبارات المتوافرة. والوحدة الخامسة والأخيرة تركز على الأسئلة والقضايا التي تظهر مرتبطة باستخدام درجات الاختبار وتفسيرها. وقد تم طرح مجموعة من التمارين في نهاية كل فصل لتقييم القارئ واختبار معرفته للمحتوى المطروح في الفصل.

وكل وحدة في الكتاب مُنَظَّمة بحيث تقدمَ في البداية الأسلوب التقليدي في تناول مشكلات تطوير الإختبار، وبعد أن يصبح القارى معتاداً على هذه الأساليب وتطبيقاتها في مواقف معيارية، فإن بعض المشكلات لا يمكن حلها بسهولة بالأساليب التقليدية. وبعد تصوّر هذه المشكلات تم عرض تقنيات قياسية إضافية (حديثة في أصولها). وعادةً ما يصاحبها مناقشة لكيفية ارتباط التقنيات الحديثة بأفكار ومفاهيم تقليدية بدرجة أكبر. ومن غير المدهش أن التطورات الحديثة في نظرية القياس تعتمد أساساً على نماذج رياضية وإحصائية أكثر تعقيداً من النظرية التقليدية، ولهذا السبب فإن الفصول في كل وحدة تم ترتيبها من الأبسط إلى الأكثر تعقيداً في كل موضوع، وتم ترتيب الموضوعات المطروحة بناءاً على أسباب بيداغوجية، وليس من الضروري أن تؤدي إلى الموافقة أو التفضيل لأيٍّ الطرق الموصوفة، وبدلاً من ذلك نأمل بأن يرى القاريء أن هنالك أساليباً متعددة لحل مشكلات تطوير الإختبار، وتخضع هذه

الأساليب لافتراضات ومحددات مختلفة. ولموقف ما، فإن الاختيار الموفق للأسلوب الأمثل يتطلب فهم ذلك الموقف.

الخلاصة:

تم في هذا الفصل تعريف مصطلحات أساسية مثل الاختبار والقياس والبناء. وقد تم تفسير أهمية البناء في تصنيف السلوكات الملاحظة والمتشابهة ودورها في تطوير النظرية.

وقد تم عرض التمييز ما بين صياغة البناء وتطوير قياس للبناء نفسه. كذلك تم تحديد خمسة مشكلات تواجه العاملين في تطوير قياسات للأبنية النفسية، هي:

1- لا توجد طريقة واحدة لتعريف البناء النفسي ومتفق عليها عالمياً.

2- تعتمد القياسات النفسية على عينات محددة من السلوك.

3- ينتج عن معاينة السلوك أخطاء قياس دائماً.

4- عدم وجود وحدات قياس محددة بالضبط.

5- يجب أن ينشأ عن القياسات علاقات إجرائية مع متغيرات أخرى كي يصبح لها معنى.

وتعد نظرية القياس فرعاً من فروع المعرفة المكرّس لدراسة كيفية تأثير هذه المشكلات على القياسات النفسية، واقتراح أساليب للتغلب على هذه المشكلات أو تقليلها. وتاريخياً، فإن تطور نظرية القياس تزامن مع تطور علم النفس نفسه، ونظرية القياس الحديثة تأثرت بالمساهمات الأولى لعلماء الإجتماع الألمان والبريطانيين والفرنسيين. وتلعب نظرية القياس دوراً مهماً في العمليات الكلية لمنهجية البحث، وذلك بطرحها أساليب عامة في أجرأة أو قياس المتغيرات التي نهتم بها، وذلك من أجل اختبار حساسية ودقة طرائق القياس التي تم تطويرها.

التمارين:

1/ بَيِّن في أي من الأنشطة الآتية استخدم القياس، الاختبار أو كليهما:

أ- طلب المعلم من التلاميذ إكمال حل خمسة مسائل في نهاية الوحدة الدراسية وجمع

أوراق الحل.

ب- سجل تلميذ في علم النفس التجريبي الوقت اللازم للفأر حتى يضغط على مسطرة عند سماعه صوت الـ buzzer.

جـ- لاحظ عالم نفس اجتماعي التفاعل بين أفراد الجنس الواحد وبين أفراد الجنسين المختلفين وذلك بأن يطلب منهم الجلوس في مقعد مزدوج في الباص العام في وقت سابق.

د- وزع معلم الصف الأول الإبتدائي الطلبة بناءً على عدد الحروف الصحيحة التي كتبوها وذلك بعد حصوله على نسخة من كتابة الأحرف الهجائية لكل منهم.

هـ- أجرى باحث مقابلة تليفونية مع المختبرين وطلب من كل واحد منهم الإجابة عل سلسلة أسئلة بـ (نعم) أو (لا) وذلك لعدد من المنتجات التجارية.

و- أجرى الباحث جدولة لعدد الاستجابات نعم وذلك لكل منتج سُئِل عنه المستجوبون.

2/ افترض توصيف الفقرات على النحو المبين أدناه، وقرر ما إذا كانت الفقرات تقيس أقصى أداء أو الأداء المميز، أو الأداء الملاحظ، وذلك لكل من:

أ- فقرات تطلب من المفحوص حل مسألة كتابية باستخدام دوال افتراضية تم تعلمها بمبحث الفيزياء.

ب- وضع إشارات على الفقرات من قبل المعلم وذلك بعد ملاحظته لأداء الطلبة في تجارب بمختبر الفيزياء تمت تحت إشراف المعلم.

جـ- فقرات تطلب من الطلبة وضع إشارة أمام كل نشاط يستمتعون به وذلك ضمن قائمة بأنشطة في درس فيزياء.

د- تتطلب فقرات الإبداع العلمي من الطلبة قراءة وصف طبيعة الظاهرة وتوليد أكبر عدد ممكن من التفسيرات المنطقية لها.

3/ اقترحت طريقة جديدة لمساعدة البالغين في اكتساب الطلاقة في لغة ثانية. وأراد الباحث تحديد ما إذا كانت الطريقة الجديدة أكثر كفاءة من الطرقة التقليدية في التدريس. ولقياس الأداء في الطلاقة باللغة الثانية، قرر الباحث عرض سلسلة مقالات مكتوبة باللغة الثانية وطلب منهم بعدها الإجابة عن الأسئلة المتعلقة بالمحتوى باتباع ما يأتي: استماع الأفراد إلى المقالات المسجلة في اللغة الثانية والإجابة عن الأسئلة المتعلقة

بالمحتوى- ما مشكلات القياس الأساسية لهذه الطريقة؟

4/ في المواقف الإختبارية التطبيقية الآتية، بين أيُّ سيكولوجيي القرن التاسع عشر كان لهم الفضل في هذا الموقف:

أ) اختيار فقرات إستبانة القلق الرياضي باختيار الفقرات التي وافق عليها المتخصصون بالرياضيات وليس طلبة الصفوف العلاجية.

ب) معلم سيطبق في الحال اختبار قرأ ما يأتي من دليل الاختبار «الرجاء قراءة التعليمات المبينة أعلى الصفحة قراءة صامتة وفي الوقت نفسه إقرأها بتأني.

جـ) استخدام إختبار -الشخصية لأنه ثبت أن درجات هذا الإختبار وضعت المستخدمين بالترتيب نفسه الذي قدمه المشرفين على العمل تقريباً بعد ثلاثة أشهر من العمل الوظيفي.

5/ إقرأ الوصف الآتي للدراسة الإفتراضية: «يريد مقوم تربوي مقارنة آثار منهاجي العلوم للصف الثامن. تتطلب الطريقة (أ) من المعلمين استخدام مقرر معين وكتاب عمل (Work book) للتلاميذ، بينما تتطلب الطريقة (B) من المعلمين تطوير خطط خاصة بالدروس حول سلسلة من التجارب المخبرية. وفي بداية السنة اختير (12) صفاً من مدارس المقاطعة بحيث تمت مزاوجتها مع عدد سنوات خبرة معلميهم، اختير منها (6) صفوف عشوائياً للطريقة (أ) و (6) الأخرى للطريقة (ب). وفي نهاية السنة اقترح الباحث (50) فقرة تقيس المبادىء العامة وتطبيقاتها اليومية واختبر الطلبة جميعهم بها. وكانت كل فقرة بأربعة بدائل، أعطيت درجة واحدة لكل إجابة صحيحة، وحسب الباحث معامل الإرتباط بين درجات نسبة الذكاء للطلبة ودرجاتهم في اختبار العلوم. وحسب الباحث أيضاً متوسط عدد الفقرات التي أجابت عنها كل مجموعة إجابة صحيحة، وقارنها باختبار الدلالة الإحصائية.

خذ بعين الإعتبار الخطوات الآتية في عملية البحث (صياغة الشكلة، تصميم البحث، المعاينة، تطوير الآداة، تحليل الاستدلال الإحصائي، زاوج بين هذه وأحد الأنشطة الآتية:

أ) أختيرت ستة صفوف عشوائياً للطريقة (أ).

ب) اقترح الباحث (50) فقرة تقيس المبادى العلمية.

جـ) مقارنة متوسط عدد الاستجابات الصحيحة للمجموعتين.

د) اختيار (12) صف من مدارس المقاطعة.

6/ لأي من الفعاليات الموصوفة في التمرين الخامس فرع (أ،ب،جـ، د) تُعد معرفة نظرية القياس أكثر أهمية؟

الفصل الثاني

2

المفاهيم الاحصائية لنظرية القياس

الفصل الثاني

المفاهيم الاحصائية لنظرية القياس

يستخدم الاحصاء في العلوم التربوية والاجتماعية لهدفين أساسيين، هما الوصف والاستدلال. فالإحصاء الوصفي يلخص أو يصف توزيع قيم مجموعة الملاحظات. والإحصاء الاستدلالي يختبر الفرضيات حول ما إذا كانت عينة معينة من الملاحظات المأخوذة من مجتمع له توزيع معين لخصائصه. ودراسة الإحصاء الاستدلالي يعد فرع من المعرفة له خصوصية ليست هدفاً لهذا الفصل. والهدف الرئيسي لهذا الفصل هو تزويد القارئ بمراجعة للإحصاء الوصفي الأساسي الذي يفيد في تفسير درجات الاختبار وتقييم نوعيته. والقاريء الذي له خلفية في الاحصاء التطبيقي يمكنه تجاوز هذا الفصل.

درجات الاختبار على أنها متغيرات وصفية لمجتمع محدود:

افترض أن معلم لديه في الصف ثلاثون طالباً أو محلل نفسي لديه (150) طالباً من ذوي صعوبات التعلم الذين أُحيلوا للمعالجة لمدة سنة. ومع أن كل من المعلم والمحلل النفسي مهتم بخصائص الأفراد جميعهم إلا أنهم يميلون إلى تلخيص بعضاً من خصائص المجموعة لاستخدامها في التخطيط لبرنامج أو للتقويم. وإحدى الطرائق الملائمة والفعالة في تلخيص المعلومات هي الإحصاء الوصفي.

وفي دراسة مجتمع ما فإن الخصائص الوصفية للأفراد قد تكون ثابتة أو متغيرة. والقيمة الثابتة تكون نفسها لأفراد المجتمع جميعهم، وفي تلخيص بيانات طلاب صف معين، فالمعلم لا يزعج نفسه بتسجيل البيانات الثابتة (مثل: كل طالب له رأس واحد، فعدد الرؤوس ثابت وله القيمة (1)) والقيمة المتغيرة هي خاصية قد تتغير من فرد لآخر في المجتمع، والمتغير قد يكون نوعي أو كمي. فلون العين متغير نوعي والوزن متغير كمي ودرجة الاختبار متغير كمي. افترض أن معلم أعد اختبار في الأحياء مؤلف من (10 فقرات وتقدم إليه (50) طالب، ويحصل المفحوص على نقطة واحدة لكل فقرة يجيب عليها إجابة صحيحة، ولا يُعطى أي درجة جزئية، لذا فإن درجات الاختبار تتراوح بين (صفر و 10). فإن أخذنا بعين الاعتبار الدرجات الملاحظة فقط فإن درجة الاختبار تكون متغير متقطع وذلك لأن القيم الممكنة محددة بدرجات أو قيم معينة على خط الإعداد الحقيقي. ففي هذه الحالة لا إمكانية للحصول على

الدرجة (2.5) أو (3.75). إضافة إلى ذلك نعرف أن الصف الدراسي فيه عدد محدد من الطلبة، وبافتراض درجة معينة لكل طالب، يتوافر لدينا عدد محدد من الملاحظات. والتفسير الاحصائي المترتب على ذلك يكون لموقف يتألف من عدد محدد من الملاحظات لمتغير متقطع. واصطلاحاً فإن القيم المحسوبة لمجتمع الملاحظات نرمز لها باستخدام الحروف اللاتينية، ونطلق عليها اسم بارامترات أو معالم. وسنعتمد هذا الاصطلاح في تعريف المفاهيم الأساسية المهمة المستخدمة في وصف توزيع الدرجات في نظرية القياس. وفقط عند ذكر ملاحظات معينة فإن مفاهيم الوحدة الأولى تم تعريفها لمجتمع الملاحظات.

الجداول التكرارية والمنحنيات

لنعد إلى معلم الأحياء الذي أعد الاختبار المؤلف من (10) فقرات ولخمسين طالباً. فإذا أراد المعلم تلخيص أداء المجموعة الكلية، فإن إحدى الطرائق المهمة والحساسة هي تحديد التوزيع التكراري للمتغير. ويستخدم التوزيع التكراري لتخصيص القيم التي تظهر للمتغير وكذلك كيفية ملاحظة هذه القيم. وتدون التوزيعات التكرارية غالباً على شكل جدول أو منحنى بياني، ولنفترض أن x تمثل المتغير (درجة اختبار الأحياء)، ويوضح الجدول (2-1) الجدول التكراري لدرجات المفحوصين وعددهم (50) على اختبار الاحياء المؤلف من (10) فقرات. وتتراوح الدرجات المكنة بين (صفر و 10)، ومرتبة تنازلياً في العمود الأول أقصى اليمين، ويبين العمود المرمز له ب f(x) عدد المفحوصين الذين حصلوا على كل درجة أو بمعنى آخر هو تكرارات كل قيمة (x). ويشير العمود الثالث إلى التكرار التراكمي لكل درجة (x)،

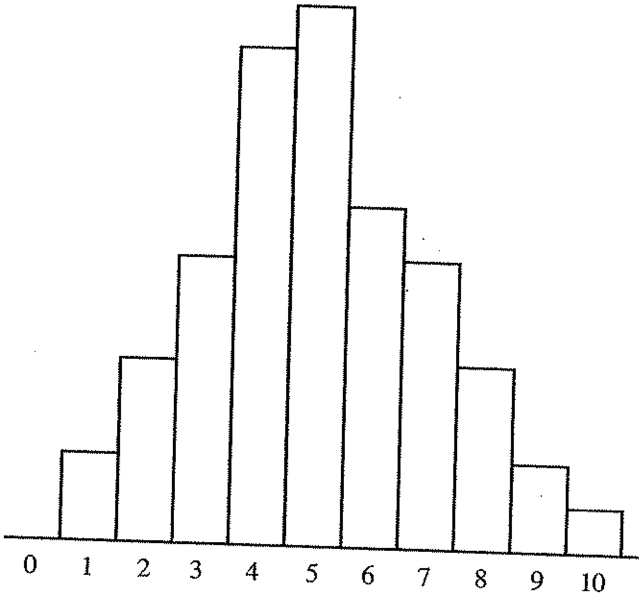
جدول (2-1): الجدول التكراري لدرجات المفحوصين على الاختبار المؤلف من 10 فقرات.

Cp (x)	Pc(x)	cf(x)	f(x)	x
1.0	.02	50	1	10
.98	.04	49	2	9
.94	.06	47	3	8
.88	.10	44	5	7
.78	.12	39	6	6
.66	.24	33	12	5
.42	.22	21	11	4
.20	.10	10	5	3
.10	.06	5	3	2
0.04	.04	2	2	1
0.00	.00	0	0	0

والتكرار التراكمي هو عدد المفحوصين الكلي الذين تكون درجاتهم تساوي القيمة (x) أو أقل منها. لهذا فإن cf(x) لـ x = 3 تكون (5 + 3 + 2 + 0 = 10)، أي المجموع الكلي لتكرارات الدرجات (0,1,2,3). ويشير العمود الرابع إلى نسبة الطلبة الذين حصلوا على درجة معينة x ويرمز لها بـ p(x)، ويطلق عليها اسم «التكرار النسبي». لهذا فإن احتمالية حصول الطالب عشوائياً على الدرجة (10) من مجموع الطلبة الكلي يكون (1/50) أو (.02). ويشير العمود الأخير إلى الاحتمال التراكمي لكل درجة (x)، وهو النسبة المئوية للذين حصلوا على الدرجة (x) أو أي درجة أقل منها.

ويمثل الشكل (2-1) مدرج تكراري لبيانات جدول (2-1)، ويُمثل المحور الأفقي الدرجات (x)، بينما التكرار النسبي لقيم (x) تمثل على المحود الرأسي. ومن هذا التمثيل نرى أن الدرجات (4 , 5 , 6) تظهر بتكرارات أعلى من الدرجات (0 , 2 , 9) أو (10)، وذلك ببساطة من ارتفاع الخط البياني عند تلك النقطة.

إن وصف كيفية توزيع درجات الاختبار (أو الملاحظات على المتغير) غير مقيد بالجداول التكرارية والتمثيل البياني. إذ يمكن استخدام بعض الإحصاءات الوصفية لتغطي بعضاً من هذه المعلومات. وعلى الأقل فإن هناك نوعين من الإحصاءات الوصفية تستخدم في وصف توزيع الدرجات هما – مقاييس النزعة المركزية ومقاييس التشتت.



شكل (2-1): المدرج التكراري لدرجات (50) مفحوص على اختبار مؤلف من (10) فقرات

مقاييس النزعة المركزية:

تمثل مقاييس النزعة المركزية درجة الطالب النموذجية في المجموعة، وتوجد ثلاثة مقاييس شائعة الاستخدام في الاختبارات هي المتوسط والوسيط والمنوال.

المنوال: يعرف المنوال بأنه القيمة الأكثر تكراراً في مجموعة الملاحظات ولبيانات جدول (2-1) منوال التوزيع = 5، أي الدرجة التي حصل عليها أكثر عدد من المفحوصين. ومن الممكن أن يكون للتوزيع التكراري أكثر من منوال.

الوسيط: هو النقطة التي تتوسط التوزيع لمجموعة البيانات. ونظرياً هو النقطة التي يكون أدنى منها وأعلى منها 50% من الدرجات. ويطابق الرتبة المئينية الخمسين. لتوزيع الدرجات. ولبيانات جدول (2-1) يقع الوسط بين الدرجتين (4, 5) أو الصيغة الحسابية لتحديد قيمة الوسيط بالضبط مبينة في فصل 19).

المتوسط: هو معدل قيم الدرجات أو البيانات جميعها، ويرمز له بالرمز μ، ويحسب من الصيغة:

$$\sum_{i=1}^{N} X_i \quad = \quad \mu \qquad \text{................ (1-2)}$$

حيث ترمز X_i إلى درجة الفرد i، وتتألف المجموعة من أفراد عددهم N، وإشارة المجموع $\sum_{i=1}^{N}$ عبارة عن رمز جبري شائع الاستخدام يعني مجموع القيم x بدءاً من الدرجة الأولى وإنتهاءاً بالدرجة N. وفي حساب المتوسط والاحصائيات الأخرى الموصوفة في هذا الكتاب على القارىء افتراض أن إشارة المجموع تطبق على البيانات كلها عندما يكون الحد الأدنى والحد الأعلى للبيانات غير محدد. لذا لمجموعة البيانات: 160 , 155 , 145 , 140 , 150 يكون

$$750 = 150 + 140 + 145 + 155 + 160 = \sum x_i$$

$$150 = \frac{750}{5} = \frac{\sum x}{N} = \mu$$

ويطلق على متوسط التوزيع أيضاً «القيمة المتوقعة لـ x».

وعندما تكون البيانات منظمة في جدول تكراري مثل جدول (2-1)، فإن المتوسط الحسابي يحسب باستخدام الصيغة:

$$\mu = \frac{\sum f_i X_i}{N} \quad \text{.....................} \quad 2-2$$

حيث ترمز X_i إلى قيم X_i المحتملة جميعها، و f_i إلى تكرار القيمة. ولبيانات جدول (2-1) فإن المتوسط يساوي

$$\mu = \frac{(0)0 + (1)2 + (2)3 + (3)5 + (4)11 + (5)12 + (6)6 + (7)5 + (8)3 + (9)2 + (10)1}{50}$$

$$\mu = \frac{250}{50} \quad \text{أو} \quad 5$$

وعبر المقاييس الثلاثة المستخدمة في وصف النزعة المركزية فإن المتوسط هو المقياس الأكثر شيوعاً واستخداماً للدرجات الاختبارية. وللعديد من الاختبارات العقلية فإن المتوسط هو الأكثر استقراراً من عينة لأخرى، وكذلك فإن متوسط العينة هو تقدير غير متحيز لمتوسط المجتمع. وهذا يعني أن المتوسطات المحسوبة من عينات عشوائية لا تكون دون التقدير أو مُغالي فيها لمتوسط المجتمع الذي أُخذت منه العينة.

مقاييس التشتت

نادراً ما تكون مقاييس النزعة المركزية مناسبة لوصف أداء مجموعة على الاختبار. ويكون هذا واضحاً إذا تصورنا أن طالب حصل على (5) درجات في اختبار الأحياء المؤلف من (10) فقرات. ومن أجل تفسير أداؤه بالنسبة لبقية الصف فإن السؤال الذي يطرحه الطالب أولاً هو: ما متوسط المجموعة؟ وعندما يعرف أن متوسط المجموعة (7) درجات فإن سؤاله الثاني من نوع: ما مدى انتشار الدرجات؟ بلغة أخرى: ما مدى التغير في الدرجات؟ فإن علم الطالب أن معظم الدرجات تتركز بين الدرجتين (6 و 8). وهنا من الأرجح أن يكون عنده بعض القلق بأن درجته منخفضة نسبياً، أما إذا عرف بأن معظم الدرجات تتوزع بين (9,2) فإنه سيكون مرتاحاً نسبياً، لأن العديد من الدرجات أدنى من درجته. ومع ذلك فإن درجته الخام ومتوسط المجموعة هو نفسه في كلتا الحالتين. إن تفسير الدرجات يتأثر بالأساس بتشتت درجات الاختبار. لذا فبالإضافة إلى مقاييس النزعة المركزية فإن مقاييس التشتت تعد ضرورية للوصف المناسب لتوزيع الدرجات. ومقاييس التشتت الأكثر شيوعاً واستخداماً هي المدى والتباين والانحراف المعياري.

ويمكن حساب مدى توزيع الدرجات من معرفة أعلى درجة وأدنى درجة في الاختبار، ومثل هذا المقياس للتشتت له خصائص عدة غير مرغوب بها. الأول أنه يتحدد بالدرجتين المتطرفتين (والتي غالباً ما تتأثر بالصدفة وخطأ القياس)، وبالتالي فإن المدى يكون غير مستقر بدرجة

كبيرة من عينة لأخرى. والثاني أن المدى لا يعطي مؤشراً عن كيفية توزيع الدرجات بين أدنى وأعلى درجة. فمن الممكن لمجموعة معينة من الدرجات أن تتراوح بين (صفر و 10)، ومن المحتمل أن تكون جميعها بين (6 و 10) ما عدا قيمة واحدة. وهنا يكون المدى مضللاً في تبيانه لدرجة تشتت درجات المجموعة.

ويعد كلاً من التباين والانحراف المعياري مقاييس تشتت تتغلب على محددات المدى. وتعتمد هذه الاحصائيات على مفهوم بسيط للدرجة المنحرفة، فلكل مفحوص درجته الخام x يمكن تحديد الدرجة المنحرفة x على أنها:

$$X - \mu = x \qquad \text{3 - 2 ....}$$

وعلى هذا فإن الدرجة المنحرفة تكون عبارة عن المسافة بين درجة المفحوص ومتوسط المجموعة، فإن كانت الدرجة الخام أقل من المتوسط فإن الدرجة المنحرفة تكون سالبة، وأما إن كانت الدرجة الخام أعلى من المتوسط فإن الدرجة المنحرفة تكون موجبة. ويمكن الحصول على تباين درجات المجموعة بإيجاد متوسط مربع الدرجات المنحرفة. وهذا يحدد بوساطة العلاقة:

$$\sigma_x^2 = \frac{\sum (x-\mu)^2}{N} = \frac{\sum x^2}{N} \qquad \text{4 - 2 ....}$$

والانحراف المعياري هو الجذر التربيعي للتباين، أي

$$\sigma_x = \sqrt{\frac{\sum (x-\mu)^2}{N}} = \sqrt{\frac{\sum x^2}{N}} \qquad \text{5 - 2..................}$$

ولتوضيح حساب كل من التباين والانحراف المعياري، دعنا نعود إلى مثال الدرجات الخمسة (160 , 155 , 145 , 140 , 150). تذكر أن المتوسط = 150. لذا فإن الدرجات المنحرفة المناظرة للدرجات الخام هي (10 , 5 , −5 , −10 , 0) على التوالي. وتطبيق المعادلة (4–2)، فإن التباين هو:

$$\sigma_x^2 = \frac{(10)^2 + (5)^2 + (5-)^2 + (10-)^2 + (0)^2}{5} = 50$$

وبأخذ الجذر التربيعي للتباين نحصل على الانحراف المعياري لتوزيع الدرجات.

$$\sigma_x^2 = \frac{\sum x^2}{N} = 7.07$$

وعندما تكون الدرجات منظمة في جدول تكراري، فإنه يمكن حساب التباين من الصيغة:

$$\sigma_x^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{N} \qquad 6-2$$

حيث ترمز x_i إلى الدرجة المنحرفة (x_i)، و f_i إلى تكرار الدرجة و k إلى عدد القيم الممكنة لـ x. ولبيانات جدول (2-1) يحسب التباين على النحو الآتي:

$$\sigma_x^2 = \frac{1(5)^2 + 2(4)^2 + 3(3)^2 + 5(2)^2 + 6(1)^2 + 12(0)^2 + 1(-1)^2 + 5(-2)^2 + 3(-3)^2 + 2(-4)^2 + 0(-5)^2 = 500}{50}$$

$$= \frac{200}{50} = 4$$

ومع أنّ حساب الانحراف المعياري سهل نسبياً، إلا أن طلبة القياس غير متيقنين من معناه. فكيف يمكن تفسير هذا الاحصائي في المواقف التطبيقية؟ الأول: تذكر أن الإنحراف المعياري والتباين يعتمدان على الدرجات المنحرفة. فكما كانت القيم المطلقة للدرجات المنحرفة أكبر كلما كان الانحراف المعياري والتباين أكبر. افترض أن مجموعة طلبة مدرسة ثانوية أجابوا على مقياس اتجاهات نحو مساويء المخدرات. وعند جدولة الدرجات للمجموعة أ كان متوسطها (45) وانحرافها المعياري (10)، والمجموعة ب كان متوسطها (45) وانحرافها المعياري (17). ويمكننا الاستنتاج من الانحراف المعياري الأكبر للمجموعة ب بأن المجموعة ب أكثر تشتتاً أو أقل تجانساً في اتجاهاتهم من المجموعة أ. نحو مساويء المخدرات.

وصف الأداء الفردي بالدرجات الزائية:

مع أن المتوسط والانحراف المعياري هما الأكثر استخداماً في وصف خصائص أداء المجموعة فإنهما يستخدمان أيضاً في إيجاد تحويلات للدرجات، والذي يجعل درجات الاختبار قابلة للتفسير بشكل أفضل. والتحويل الأكثر شيوعاً للدرجات هو الدرجة الزائية، وتحسب باستخدام الصيغة:

$$Z = \frac{X - \mu}{\sigma} \qquad (7-2) \text{......................}$$

حيث ترمز X إلى الدرجة الخام، و μ إلى متوسط المجموعة، و σ إلى الإنحراف المعياري. وتشير الدرجة الزائية إلى عدد الانحرافات المعيارية التي تقع أعلى أو أدنى من الدرجة X، في حين أن قيمة الدرجة الزائية تشير إلى المسافة بين الدرجة الخام ومتوسط المجموعة بوحدات الانحراف المعياري، وتشير إشارة الدرجة الزائية إلى ما إذا كانت الدرجة أعلى أو أدنى من المتوسط. وهذا الاستخدام في ملف الاختبار المبين في جدول (2-2). وبعد فحص الدرجات الخام لهذا الطالب فإن المرشد التربوي سيحاول ارشاد الطالب إلى التخصص في الرياضيات أو العلوم أو الإنسانيات. وإذا اختبر الفاحص الدرجات الزائية التي تأخذ بعين الاعتبار متوسطات وتباينات الاختبارات. ومن الواضح أن هذا الطالب سيجد أن الفنون أكثر توافقاً مع استعداده. وفي تفسير الدرجات الزائية من الضروري أن نضع في اعتبارنا أن متوسط التوزيع = صفر وانحرافه المعياري يساوي واحد.

جدول (2-2): ملف الدرجات الخام والدرجات الزائية المناظرة لمفحوص واحد

الاختبار الفرعي أ	الدرجة الخام	الدرجة الزائية
الرياضيات	102	,31-
فنون اللغة	80	1.25
العلوم الطبيعية	115	.10
الإنسانيات	95	.83

أ- المعلومات الآتية المتوافرة لكل اختبار فرعي

اختبار الرياضيات	150 فقرة	110 = μ	26 = σ
اختبار اللغة	110 فقرات	70 = μ	8 = σ
اختبار العلوم الطبيعية	125 فقرة	112 = μ	30 = σ
اختبار الإنسانيات	140 فقرة	90 = μ	6 = σ

حسبت المتوسطات والانحرافات المعيارية لمجتمع المفحوصين نفسه

المجتمعات غير المحددة (اللانهائية) والمتغيرات المتصلة:

أحياناً، نأمل في نظرية القياس أن نفكر في توزيع متغير لا نهائي (غير محدود). ففي مثل هذه المواقف لا تستخدم المعادلة (2-1) لحساب المتوسط ولا المعادلة (2-5) للإنحراف المعياري. وتوجد صيغة مكافئة للمعادلة (2-1) يمكن تطبيقها بالتساوي للمجتمعات المحدودة وغير المحدودة من البيانات هي:

$$\mu = \sum P_k x_k \qquad 8 - 2$$

حيث ترمز X_k إلى قيمة x للفرد k و p_k إلى احتمالية ظهور القيمة x. لاحظ أنه عند تطبيق هذه الصيغة لمجتمع محدد من البيانات (مثل تلك في جدول 2-1) فإنها ستتجاوز القيمة الناتجة عن تطبيق المعادلة (2-1).

بصورة مشابهة، يتحدد الانحراف المعياري لمجتمع غير محدود من البيانات بالصيغة:

$$\sigma_x = \sqrt{\sum P_k (x_k - \mu)^2} \qquad 9 - 2$$

ومرة أخرى عند تطبيق هذه المعادلة على مجموعة محدودة من البيانات أو الدرجات فإننا سنحصل على نتائج مشابهة لتلك الناتجة عن المعادلة (2-5).

ومن المهم ملاحظة أن الإحصاء يشير أحياناً إلى خصائص التوزيع التكراري على أنه عزوم، المتوسط كما هو محدد في المعادلة (2-8) يسمى العزم الأول لأصل التوزيع، والانحراف المعياري كما هو محدد في المعادلة (2-9) يسمى العزم الثاني حول متوسط التوزيع.

ومع ذلك فإن درجات اختبار مثل الاختبارات الصفية واختبارات الذكاء ومقاييس الاتجاه تفترض مجموعة من الدرجات أو القيم المتقطعة؛ والسمة المتضمنة في الأداء على الاختبار تعد مع ذلك متغير متصل. افترض أن فقرة واحدة تتطلب من المفحوص تهجئة كلمة (armament). افترض أن المحوصين (أ ، ب ، جـ) لفظوها جميعاً بصورة خاطئة، وحصلوا جميعاً على درجة صفر. ومع ذلك فإن القدرة المتضمنة في عملية التّهجئة للكلمة قد تكون مختلفة عند الثلاثة. فالمفحوص أ قد لا يعرف كيف يبدأ تهجئة الكلمة، والمفحوص ب قد يكون بدأ ...a-r-m ولكنه غير قادر على تكملة الكلمة بصورة صحيحة، والمفحوص جـ قد يكون تهجأ الكلمة على أنها a-r-m-i-ment ويكون أخطأ بحرف واحد فقط. ومن الواضح أنه يوجد تقسيم أكثر دقة لقدرة التّهجئة من الذي تعكسه درجة الاختبار. كذلك وبغض النظر عن عدد فقرات الاختبار وكيف تم تطوير قواعد التصحيح من الممكن أن نتصور أن المفحوص له قدرة

أعلى أو أدنى من تلك التي تعكسها الدرجة الملاحظة، فإن كانت الدرجات الملاحظة , 41 , 42) (40 وهكذا، فمن الممكن أن نتخيل بأن مفحوص ما له الدرجة (40.5) على متصل السمة نفسها، وإذا وسعنا التدريج ليقيس (40.1 , 40.2 , 40.3) سيبقى هنالك إمكانية لنتخيل بأن هنالك مفحوص له الدرجة (40.15) على متصل السمة نفسه وهكذا، والمتغير المتصل له قيم ممكنة يمكن تحديدها باستمرار بفترات أصغر وأصغر. فبين الحدود العليا والحدود الدنيا يفترض للمتغير المتصل أية قيمة على خط الأعداد الحقيقية. وعندما تقابلنا بيانات متغير متصل فإن الصيغ المبينة لكل من المتوسط والتباين والانحراف المعياري لا تكون مناسبة. فللبيانات المتصلة لا يمكن إجراء عملية الجمع. والصيغ المناسبة لتحديد العزوم (المتوسط والانحراف المعياري) لمتغير ذي توزيع متصل يتطلب عمليات حساب تكامل، ولن نعرض هذه الحسابات هنا. ومع ذلك فإن على الطالب أن يميز أن بعضاً من مفاهيم نظرية القياس تعتمد على فكرة المتغير المتصل وبالتالي فإنها تتحدد باستخدام الاقترانات والعمليات المستمدة من الرياضيات.

التوزيع الطبيعي

أحد المفاهيم الأساسية لنظرية القياس هو التوزيع الطبيعي، وهو توزيع نظري لمتغير متصل مقاس لمجتمع لا نهائي. واهتم الرياضيّون بدءاً من القرن التاسع عشر أمثال كوتيلت وجالتون بإيجاد التوزيعات التكرارية لعدد كبير من الملاحظات (البيانات) على متغير للسمات الإنسانية تقترب من شكل الجرس والذي نطلق عليه اسم المنحنى الطبيعي. وهذا صحيح حتى عندما تقاس السمات على تدريج متقطع ولعينة محدودة. ومع أن بعضاً من بيانات التوزيع قد تمثل شكل المنحنى الطبيعي فإن التوزيع الطبيعي للبيانات الحقيقية المستخلصة من عدد محدود من القيم المتقطعة لا يمكن أن يكون أكثر من تقريب لهذا المنحنى النظري.

إن المنحنى الطبيعي النظري (شكل 2-2أ) عبارة عن تمثيل بياني لدالة رياضية اشتقها الفرنسي كارل جاوس. ويتحدد شكل المنحنى الطبيعي بالمعادلة:

$$y = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \qquad 10 - 2$$

وتمثل y في هذه المعادلة ارتفاع المنحنى فوق قيمة معينة x ، و σ تمثل الانحراف المعياري لتوزيع درجات x، وتمثل μ المتوسط. والقيم الثابتة في هذه المعادلة هي π وتساوي 3.1416 و $e \approx 2.718$ وهو أساس اللوغاريتم الطبيعي. ولأن قيم المتوسط والانحراف المعياري تكون

مختلفة لكل مجموعة من الملاحظات فإن المعادلة (2-10) لا تعرف منحنى واحد بل عائلة من المنحنيات التي تشترك في خصائص معينة.

ومع أن المعادلة (2-10) قد تظهر منفردة، فإن المثال التطبيقي يفيد في حل بعضاً من الغموض في هذه المعادلة المعقدة. في البداية لاحظ أن $(x - \mu)^2/\sigma^2$ هي Z^2 لأي قيمة x، ولاحظ أن الكمية $\frac{1}{\sqrt{2\pi\sigma}}$ هي نفسها لجميع قيم x في التوزيع. افترض أننا نريد رسم التوزيع الطبيعي الذي متوسطه = 50 وانحرافه المعياري 10، في هذه الحالة فإن $\frac{1}{\sqrt{2\pi\sigma}}$ تساوي (.04) لقيم x جميعها في التوزيع. دعنا نفترض ولإهداف التوضيح تحديد قيمة y للدرجة x = 50. في هذه الحالة فإن الدرجة المنحرفة x = x - μ = صفر. وبالتالي فإن z = صفر. لهذا فإن

$$y = 0.04 \, (e)^{-\frac{1}{2}(\text{صفر})^2}$$

وتذكر أن أي رقم مرفوع للأس صفر=1، يمكن كتابته على شكل كسر ذي أس موجب، لذا فإن:

$$y = 0.04\,(e)^{-2} \quad / \quad y = 0.04\ e^{-2}$$

وحيث ان اي رقم اسه سالب يمكن كتابته على شكل كسر ذو اسٍ موجب، لذا فإن

$$y = 0.04\left(\frac{1}{e2}\right) = \quad .04\left(\frac{1}{(2.718)^2}\right) = .005$$

وقيم (0.04 و 005.) التي حصلنا عليها من هذه الحلول تعطينا قيمة ارتفاع المنحنى فوق القيم (x = 050. و x = 30) على التوالي.

ويجب ملاحظة أن ارتفاع المنحنى الطبيعي عند قيم مختلفة لـ x ليست احتمالات، وذلك لأن x متغير متصل فمن المستحيل تحديد احتمال ظهور درجة معينة مثل x = 50، ومع ذلك فمن الممكن تحديد احتمالية الحصول على قيمة لفترة معلومة من الدرجات (مثل: من 49.6 إلى 50.5). وفي الرياضيات توجد عملية تسمى التكامل، وهذا التكامل يمكننامن حساب المساحة تحت المنحنى بين أي قيمتين. وهذه المساحة تعطينا احتمالية الحصول على درجة من تلك الفترة. وعند تطبيق عملية التكامل لدالة المنحنى الطبيعي فإن:

$$p(1 \le x \le 4) = \int_{2}^{4} \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{1}{2}(x-\mu)^2/6^2}\, d_x \qquad 11\text{-}2$$

حيث ترمز p إلى الاحتمالية، وترمز l و u إلى حدود الفترة الدنيا والعليا والرموز $\int$ و dx إلى عملية التكامل للتعبير الرياضي المكتوب بينهما.

وتفسر درجات الاختبار عادةً بإيجاد نسبة الحالات التي تقع أدنى (أو أعلى) من درجة معينة على افتراض التوزيع الطبيعي. ولإيجاد نسبة الحالات التي تقع دون درجة معينة فإننا نعتبر الحد الأدنى للتكامل في المعادلة (2-11) هو (-∞)، ويكون الحد الأعلى هي الدرجة التي نهتم بها، ومن ثم حساب التكامل. والناتج عن هذه العملية يكون احتمالية الحصول على درجة مساوية لـ x أو أقل منها، ومع ذلك لا يوجد من مستخدمي الاختبارات من يسعى إلى حل المعادلة (2-11) لكل درجة اختبارية. ولحسن الحظ توجد طريقة أخرى للحصول على هذه الاحتمالات دون رؤية معادلة المنحنى الطبيعي. فعند تحويل درجات التوزيع الطبيعي إلى درجات زائية فإن توزيع الدرجات الزائية يسمى التوزيع الطبيعي المعياري وله متوسط = صفر وانحراف معياري = 1.0. ويبين الملحق (أ) المساحة تحت المنحنى إلى يسار القيم الزائية للمنحنى الطبيعي المعياري، فكل قيمة تمثل احتمالية الحصول على قيمة أقل أو تساوي z التي حسبت من عملية التكامل. كيف يمكن استخدام ملحق (أ) في الحصول على نسبة المفحوصين الذين سيحصلون على درجة أقل من قيمة معينة في توزيع يقترب من التوزيع الطبيعي؟ ببساطة حوّل الدرجة الخام إلى درجة زائية وبعدها استخدم ملحق (أ) في الحصول على الاحتمالية المقابلة. مثلاً افترض أن المتوسط = 1000، والانحراف المعياري = 10 ونريد معرفة نسبة المجتمع الذين هم دون الدرجة 110. الدرجة الزائية المقابلة ($\frac{100\text{-}110}{10}$ = 1.00) وفي الملحق (أ) نجد أن المساحة النسبية تحت المنحنى إلى يسار z = 1.00 هو (.84). لهذا فإن (84%) من المفحوصين ستكون درجاتهم تساوي أو أقل من 110. وهذه الطريقة تصلح في الحساب لأن الدرجات الزائية تعد تحويل خطي للدرجات الخام، فإن كان توزيع الدرجات الخام الأصلي طبيعي تقريباً فإن الدرجات الزائية تتوزع بالطريقة نفسها، وبالتالي فإن النسبة نفسها تقع دون الدرجة الخام الملاحظة والمطابقة للدرجة الزائية.

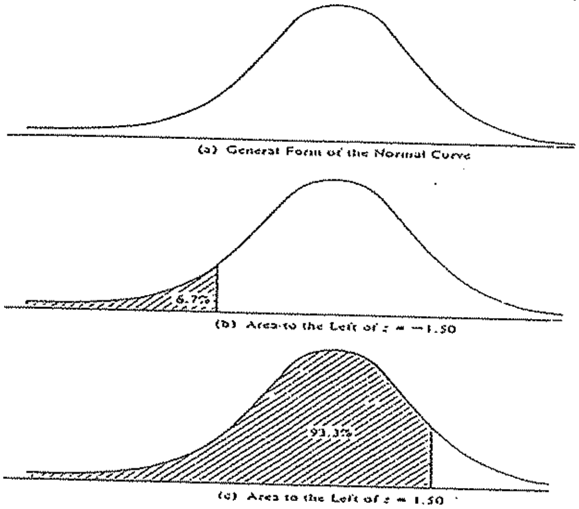
ومن المهم ملاحظة أنه توجد احتمالات مختلفة مرتبطة بالدرجة الزائية الموجبة والسالبة التي لها القيمة المطلقة نفسها. افترض على سبيل المثال أننا نريد حساب احتمالية الحصول على درجات z تساوي أو أقل من القيم (-1.50 و 1.50).

وتبين الأشكال (2 – 2ب) و (2 – 2جـ) موقع كلتا الدرجتين على التوالي، والمساحات المظللة تحت المنحنى تشير إلى احتمالية الحصول على درجة زائية تساوي هذه الدرجات أو

أقل منها. ومن هذا الشكل يمكن رؤية أن مساحة أقل بكثير تقع دون z = -1.5 من تلك التي تقع دون 1.50. وإن عدنا الآن إلى الملحق أ وموقع الدرجة الزائية (1.50) في العمود الأول فسنحصل على الاحتمالية المقابلة للدرجة z أقل أو تساوي -1.50 من العمود الثاني.

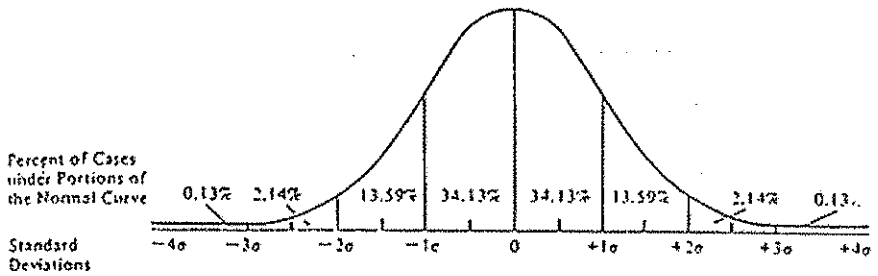
ونرى أن هذه الاحتمالية تساوي (.067). فلمجموعة مفحوصين بتوزيع طبيعي نتوقع 7% تكون درجاتهم أقل أو تساوي (-1.50) ومن العمود الثالث يمكننا معرفة أن احتمالية الحصول على درجة أقل أو تساوي (1.50) هو 0.933. وعلى هذا نتوقع أن (93%) من المفحوصين تقريباً في هذه المجموعة سيحصلون على درجة أقل من هذه النقطة.

وتشترك التوزيعات الطبيعية جميعها بمجموعة خصائص تعد مهمة في تفسير الدرجات. إحدى هذه الخصائص هو أن المتوسط والمنوال والوسيط لها القيمة نفسها – وتقع في وسط التوزيع. إضافة إلى أن كل توزيع طبيعي متماثل حول نقطة المتوسط، وذلك أن نصفي المنحنى بالضبط كل منهما مرآة للنصف الآخر. فإذا أخذنا نقطتين على المسافة نفسها أعلى وأقل من المتوسط فإن النسبة نفسها تقع أعلى وأدنى من المتوسط ضمن فترة الدرجة نفسها. وأخيراً وفي كل توزيع طبيعي وكما هو موضح في الشكل (2-3) فإن (34%) من الدرجات تقع في الفترة بين ($\mu = \mu + 1\sigma_x$) وحيث أن المنحنى منتظم فإن 34% أخرى من الدرجات تقع بين ($\mu = \mu + 1\ _x$) وعلى هذا فإن 68% من الدرجات تقع تقريباً على بعد درجة معيارية واحدة أعلى وأدنى من المتوسط. وكما هو موضح في الشكل (2-3) فإن 14% من الدرجات تقع تقريباً في الفترة ($\mu + 1\ _x$) و ($\mu + 2\ _x\sigma$) وأخيراً فإن 2% من الدرجات تقع بين $\mu + 2\sigma$ و $\mu + 3\sigma$.



شكل (2-2): المنحنى الطبيعي

دعنا نرى كيفية تطبيق هذه المعلومات في توزيع الدرجات على اختبار ذكاء معياري (مقنن) معروف، وسطه (100) وانحرافه المعياري (15). فإن اختبرت عينة طلبة من مجتمع طبيعي فإن حوالي 68% من العينة ستكون درجاتهم بين (85 و 115) وذلك لأن هذه القيم تناظر ($\mu - 1\sigma_x$) و ($\mu + 1\sigma_x$). كذلك فإن (95%) من المجموعة ستكون درجاتها بين (70 و 130) في الاختبار، إذ أن هذه القيم تناظر النقاط ($\mu - 2\sigma_x$) و ($\mu + 2\sigma_x$) وكذلك نتوقع أن أي درجة أقل من (85) تقع ضمن الـ (16%) الأدنى من التوزيع، والدرجة الأعلى من (130) تقع ضمن أعلى (2.5%) من التوزيع. وبالطبع من الممكن تحديد نسبة الحالات التي تقع أدنى أو أعلى من أي درجة معلومة في التوزيع الطبيعي، ولكن النسب المناظرة لانحرافات معيارية تبعد (1 أو 2 أو 3) عن المتوسط نستشهد بها اصطلاحاً في عرض مؤشرات لتفسير الدرجات أو في تحديد درجات قطع للدخول في البرامج المصممة لتخدم عدد محدد من الطلبة أو المسترشدين (مثل برامج التربية الخاصة).



شكل (2-3): نسبة الحالات في كل جزء من المنحنى الطبيعي

وصف العلاقة بين متغيرين

ركزنا حتى الآن على استخدام الاحصاء في وصف توزيع الدرجات على متغير واحد. وفي العديد من المواقف يهتم باني الاختبار ومستخدمة بكيفية ارتباط درجات قياسين مختلفين للأفراد نفسهم. فمثلاً، قد يهتم المحلل النفسي بمدى ارتباط درجة المريض النفسي على مقياس الكذب في قائمة MMPI مع عدد الجمل الكاذبة أثناء المقابلة. وبصورة مشابهة قد يهتم مكتب القبول الجامعي بالارتباط بين الأداء على اختبار القبول المدرسي "Scholastic

"Admission Test" بمتوسط الدرجات في نهاية الفصل الدراسي الأول، وحتى معلم المدرسة قد يهتم بمعرفة مدى ارتباط درجات اختبار الاستيعاب القرائي ومقياس مفهوم الذات الأكاديمي. وفي كل موقف من هذه المواقف يطرح مستخدم الاختبار هذا السؤال: ما هي درجة العلاقة بين الدرجات على القياسين الأول والثاني؟

وعندما نهتم بالعلاقة بين متغيرين اثنين، فإننا نريد معرفة ما إذا كانت الدرجة المنخفضة على المتغير الأول مرتبطة بالدرجات المنخفضة على المتغير الثاني، وفيما إذا كانت الدرجة العالية على المتغير الأول ترتبط عموماً بالدرجات العالية على المتغير الثاني. لاحظ أننا نهتم بالقيم النسبية عندما نتحدث عن الدرجات العالية أو المنخفضة. علاوة على ذلك فإننا نهتم فيما إذا كانت الفروق عبر الدرجات على المتغير الأول مرتبطة بفوارق مشابهة نسبياً على المتغير الثاني. بمعنى آخر علينا أن نسأل كيف يرتبط انحراف الدرجات على x (المتغير الأول) مع انحراف الدرجات على y (المتغير الثاني). وفي العديد من الحالات فإن المتغيرين (xو y) يقاسان على تدريجات مختلفة، فمثلاً تقاس درجات اختبار الاستعداد المدرسي (SAT) على تدريج مؤلف من (100) درجة على التدريج (x)، في حين أن متوسط الدرجات (GPA) والدرجات المعيارية (المتغير y) يعبر عنها بأجزاء من الدرجة الواحدة. وللملائمة بينهما يمكن تحويل درجات كل من (x و y) إلى درجات زائية. وبالتالي يصبح كل منهما على تدريج مألوف شائع متوسطه صفر وانحرافه المعياري 1.00.

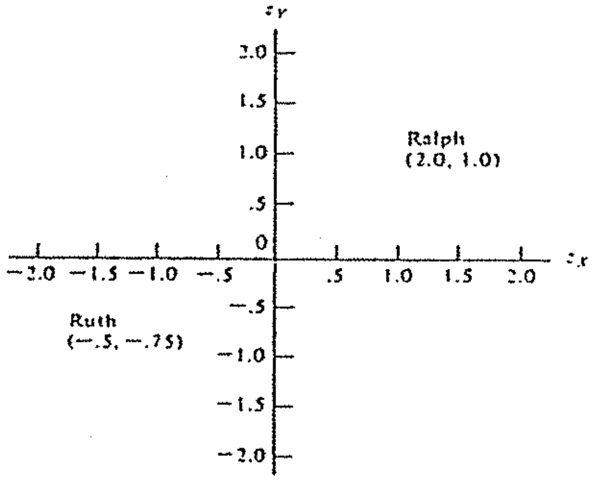
التمثيل بالأزواج المرتبة على المستوى البياني:

إحدى طرائق فهم العلاقة بين متغيرين هو التمثيل على المستوى البياني بحيث أن كل زوج من الدرجات (Zy , Zx) تمثل بنقطة واحدة كما هو مبين في الشكل (2-4). وتمثل درجات Zx على المحور الأفقي الذي يعرف بالإحداث السيني، وتمثل درجات Zy على المحور العمودي الذي يعرف بالإحداثي الصادي. وتتحدد إحداثيات كل نقطة على المنحنى بزوج من الدرجات هما Zx و Zy.

فمثلاً: درجات رالف هي Zx = 2.0 و Zy = 1.0، وهذه الدرجات مبينة في النقطة التي احداثياتها (2 , 1)، بينما درجات روث Zx = -0.5 و Zy = -0.75 تبينها نقطة أخرى. والتمثيل بالأزواج المرتبة الذي يلخصه الخط البياني هو تجميع لنقاط أفراد العينة جميعهم. والخط البياني مع العلاقة التي يمثلها هذا الخط تميز من خلال الإتجاه والشكل.

وقد يكون إتجاه العلاقة موجب أو سالب، ففي العلاقة الموجبة تميل الدرجات العالية على المتغير الأول لأن تظهر مع الدرجات العالية على المتغير الثاني، وتميل الدرجات المنخفضة على

المتغير الأول لأن تظهر مع الدرجات المنخفضة على المتغير الثاني. ويظهر الشكل (2 – 5 أ) العلاقة الموجبة بين درجات (Zx و Zy). ونتوقع عادة وجود علاقة موجبة بين متغيرات مثل اختبار الاستعداد المهني (الوظيفي) والأداء الوظيفي، وبين درجات نسبة الذكاء ودرجات اختبار الاستعداد القرائي، وبين وزن الطفل الرضيع ومستوى النمو الحس حركي. وفي العلاقة السالبة تميل الدرجات العالية على المتغير الأول لأن تظهر مع الدرجات المنخفضة على المتغير الثاني، والعكس بالعكس. ويظهر الشكل (2 – 5ب) العلاقة السلبية. ومثل هذه العلاقة يمكن أن توجد بين متغيرات مثل غياب الطلبة ووسط درجاتهم (GPA)، أو مثل الدرجات على كل من مقياس القلق وتقدير الذات.

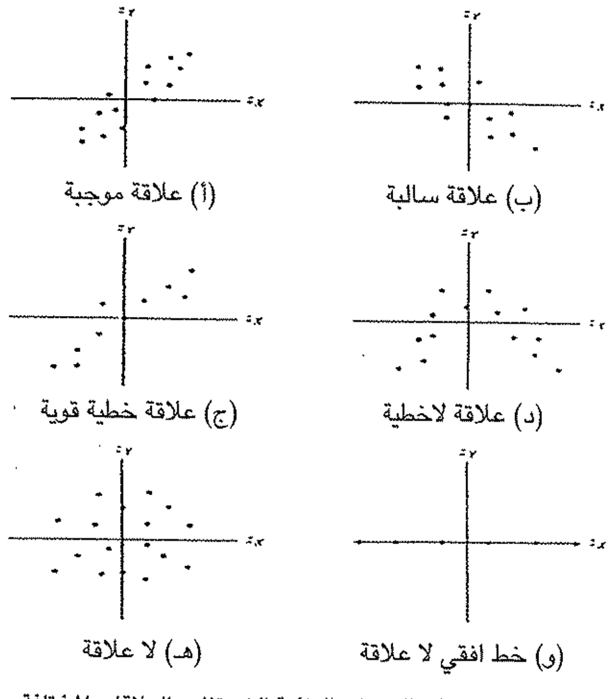


شكل (2-4): رسم للنقاط المعتمد على الدرجات الزائية لمفحوصين اثنين

ونود تمييز شكل إنتشار الخط البياني على أنه يمثل بعض الأشكال الهندسية. والشكل الأبسط الذي يمكن استخدامه في تمثيل العلاقة بين متغيرين هو الخط المستقيم. وفي وصفنا لشكل الإنتشار نهتم بمدى دقة الخط المستقيم في تمثيله للنقاط. فالشكل (2 – 5 جـ) يوضح الإنتشار عندما تقع النقاط كلها على خط مستقيم تقريباً، ومثل هذا الإنتشار يشير إلى علاقة قوية بين درجات Zx و Zy. ولأن الخط المستقيم يمكن رسمه بحيث تقع جميع النقاط قريبة جداً من الخط المستقيم. ويمكن ملاحظة مثل هذه العلاقة إذا كانت كل من (Zx و Zy) تمثل درجات على صيغتين متكافئتين لاختبار في الرياضيات متساويين في الطول والصعوبة ومتناظرين في المحتوى وتم إعدادهمافي اليوم نفسه. وشكل الانتشار البيضوي في الشكل (2 – 5أ) يبين بيانات مجموعة يمكن تمثيلها بخط مستقيم بشكل متوسط. ويمكن ملاحظة مثل هذه العلاقة بين متغيرات مثل الدافعية للتحصيل ومتوسط الدرجات مع أن ميل ذوي الدافعية

العالية لأن يحصلوا على درجات عالية في الاختبارات المدرسية، إلا أن ترتيب الأفراد على المتغيرين ليست تامة.

وتمثل الأشكال (2 – 5د) و (2 – 5هـ) حالات يمكن القول أنه لا علاقة خطية بين درجات Zx و Zy. ففي العلاقة غير الخطية (2 – 5د) يمكن تمثيل شكل الإنتشار بشكل هندسي غير الخط المستقيم بصورة أفضل. فالعلاقة بين فترة العلاج Zx ومستوى التعديل النفسي Zy يمكن أن تكون علاقة غير خطية. وشكل الإنتشار الدائري في الشكل (2–5هـ) يشير إلى علاقة ضعيفة أو لا علاقة من أي نوع بين درجات (Zy,Zx)، فلا يوجد نمط تنبؤي بين درجات Zx و Zy، فكل فرد يحصل على أي درجة Zx يظهر بأنه يمكن أن يحصل على درجة منخفضة أو متوسطة أو عالية بالتساوي على Zy. والمثال الأخير (شكل2-5و) يوضح نموذج انتشار خطي واضح للبيانات، ولكن لا علاقة بين Zx و Zy. فالعلاقة الإحصائية التي تصف مدى تغير القيم على Z المقابلة لتغير القيم على y، فإن كان هناك تغير قليل أو لا تغير في قيم المتغير x أو y فإن العلاقة تكون ضعيفة أو لا علاقة بين المتغيرين، وحالة كهذه تظهر عندما يحصل الجميع على الدرجة نفسها على أحد المتغيرين فعندما يكون الإنتشار عمودي أو أفقي تماماً فإننا نقول بأن لا علاقة بين المتغيرين.



شكل (2-5): التمثيل البياني المعتمد على الدرجات الزائية التي تظهر العلاقات المختلفة بين المتغيرين x و y

إن أشكال إنتشار مثل تلك التي في الشكل (2-5) تدعم فهم العلاقة بين المتغيرين، فمثل هذه الرسوم البيانية لا تسمح بوصف دقيق لدرجة العلاقة. فإذا نظر اثنان إلى التمثيل البياني نفسه فإنهما يستخلصان إنطباعات مختلفة عن درجة العلاقة الممثلة، وعلاوة على ذلك يمكن أن يظهر خطّان بيانيّان الشكل نفسه، وبالتالي فإن هنالك مواقع مختلفة للنقاط التي من الصعب اكتشافها من النظر إليها فقط. ولهذا السبب فمن الضروري أن يكون هنالك معامل عددي يصف درجة العلاقة الخطية بين مجموعتي بيانات المجموعة نفسها من الأفراد.

معامل ارتباط بيرسون

أحد مقاييس العلاقة الخطية بين مجموعتين من الملاحظات احصائي يعرف بإسم «معامل الارتباط». ومع أن هنالك عدة أنواع لهذا المعامل إلا أن المعامل الأكثر شيوعاً واستخداماً هو معامل ارتباط قدمه مفاهيمياً جالتون واشتقه رياضياً فيما بعد بيرسون، فعندما تمثل البيانات (y,x) بالدرجات الزائية فإن صيغة معامل الارتباط هي:

$$\rho xy = \frac{\sum Zx\, Zy}{N} \qquad \text{12 - 2 ....}$$

حيث تمثل N عدد الأفراد. ومن المهم ملاحظة أنه في حساب حاصل ضرب الدرجات الزائية فإن Zx Zy يجب أن تحسب لكل فرد قبل إجراء عملية الجمع. ويبين الجدول (2-3) الدرجات الزائية لاختبار الاستعداد ومتوسط درجات (10) أفراد مع توضيح لخطوات حساب معامل الارتباط.

وتمتد قيم معامل الإرتباط من -1.0 إلى 1.0، وتستخدم قيمة الرقم كمقياس لقوة العلاقة الخطية بين المتغيرين، وتشير الإشارة إلى الإتجاه الموجب أو السالب للعلاقة. وعندما تقترب قيمة ρxy من الصفر، فهذا يعني أنه لا علاقة خطية أو أنها ضعيفة بين المتغيرين. وعندما تقترب قيمة ρxy من 1.00 فإنها تشير إلى علاقة موجبة قوية بين المتغيرين فمثلاً تتراوح معاملات الارتباط بين اختبارات الذكاء المطبقة فردياً وجمعياً بين (0.80 إلى 0.90) ومؤشرة إلى تشابه كبير في أداء الأفراد على الاختبارين نسبياً. بالمقابل فإن معامل الارتباط بين قلق الاختبار والأداء في الاختبارات التحصيلية تتراوح عادةً بين (-20. إلى -40.) ومؤشرة إلى علاقة خطية سالبة ضعيفة.

جدول (2-3): حسابات توضيحية لمعامل ارتباط بيرسون بالاعتماد على الدرجات الزائية

ZxZy	Zy	Zx
.00	.33-	.00
.83	1.00-	.83-
2.20	1.17	1.88
.26	.50	.52
1.90	1.83	1.04
.26	.50-	.52-
.00	.00	.31
3.62	1.83-	1.98-
.14-	.33	.42-
.00	.17-	.00

$$8.93\ /\ 0.89 = \frac{8.93}{10} = \frac{\sum Zx\,Zy}{N} = \rho xy\ (= \sum Zx\,Zy\ )$$

ويمكن حساب معامل الارتباط من الدرجات الخام أو الدرجات الزائية. تذكر أن

$$Zx = \frac{x - \mu x}{\sigma x}$$

$$Zy = \frac{y - \mu y}{\sigma y}$$

فإذا عوضنا هذه الكميات Zx و Zy في المعادلة (2-12) سنصل إلى الصيغة الآتية والمعبر عنها بالدرجات الخام:

$$\rho xy = \frac{\sum (x-\mu x)(My-\mu y)}{N\sigma x\,\sigma y} \quad \text{13 -2......................}$$

لاحظ أن الكميات (Mx-μx) و (My-μy) يمكن كتابتها كدرجات منحرفة، وعلى هذا فإن التعبير الثالث لمعامل الارتباط هو:

$$\rho xy = \frac{\sum x\,y}{N\sigma x\,\sigma y} \qquad \text{14-2} \; ....$$

وثانية، من المهم ملاحظة أنه يجب حساب حاصل ضرب الدرجات المنحرفة ثم جمعها للأفراد جميعهم، ومع ذلك فقد تم عرض ثلاثة صيغ لمعامل الارتباط تعطي جميعها القيمة العددية نفسها. وسنجد أن صيغ الدرجات الزائية والدرجات المنحرفة مفيدة خاصة في الاشتقاقات العديدة المعروضة في الفصول التالية. ويوضح الجدول (2-4) حساب معامل الارتباط من الدرجات الخام باستخدام المعادلة (2-14) (لاحظ أن الدرجات الخام المقابلة للدرجات الزائية على اختبار الاستعداد والمتوسط للدرجات نفسها المبينة في جدول (2-2).

تفسير معاملات الارتباط

لا مناقشة لمعامل الارتباط يمكن إكمالها دون مناقشة الاعتبارات التي تؤخذ بالاعتبار عند تفسيرها. إحدى هذه الاعتبارات هي ما إذا كانت قيمة معامل الارتباط كبير إلى درجة كافية ليؤشر إلى علاقة مهمة أو ذات معنى بين متغيرين أثنين. على الأقل يوجد محكين شائعي الاستخدام في الحكم على أهمية الارتباط الإحصائي:

1) هل معامل الارتباط يختلف بدلالة عن الصفر؟

2) ما النسبة المئوية للتباين على x والمشترك مع التباين على y؟

لتطبيق المحك الأول على الباحث أن يأخذ بعين الاعتبار إمكانية كون العينة مسحوبة من مجتمع له $\rho xy = 0.00$ ولاختبار هذه الإمكانية على الباحث أن يؤسس فترة من القيم التي تحتوي على (95%) من معاملات الارتباط الناتجة عن (100) عينة حجمها N أخذت من مجتمع له قيمة $\rho xy = 0.00$. وهذا المدى يكون منتظماً حول الصفر، فإن كان الإرتباط المحسوب من العينة الفعلية يقع خارج هذا المدى فإن الباحث سيستنتج أن الارتباط الملاحظ يختلف عن الصفر وبدلالة (وتذكر أن الباحث يكون واثقاً بنسبة 95% فقط لاستنتاجه). فإن كان الارتباط المحسوب للعينة الفعلية يقع ضمن المدى المشار إليه، فإن الباحث لن يعدل استنتاجه بأن الإرتباط الملاحظ يختلف اختلافاً دالاً عن الصفر. ولتكوين الفترة الضرورية حول ρxy = صفر على الباحث أن يفترض بأن القيم للعينات الافتراضية المئة تتوزع اعتدالياً حول ρxy = صفر. تذكر أنه وفي التوزيع الطبيعي يقع (95%) من الحالات تقريباً على بُعد انحرافين معياريين دون المتوسط وأعلى منه. وقد قدم لنا ماغنسون الصيغة أدناه لتقريب الإنحراف المعياري لمعامل الارتباط حول ρxy = صفر، وكالآتي:

$$\sigma_p = \frac{1}{\sqrt{N-1}} \qquad 15 - 2$$

حيث تمثل N حجم العينة، وعند تطبيق الصيغة على N = 50، ينتج:

$$\sigma_p = \frac{1}{\sqrt{1-5}} = 0.14$$

جدول (2-4): حساب توضيحي لمعامل ارتباط بيرسون المعتمد على الدرجات الخام

xy	(y-y) أو x	(x-x) أو x	y	X
0.0	.2-	0	2.1	100
4.8	.6-	8-	1.7	92
12.6	.7	18	3.0	118
1.5	.3	5	2.6	105
11.0	1.1	10	3.4	110
1.5	.3-	5-	2.0	95
0.0	0	3	2.3	103
20.9	1.1-	19-	1.2	81
.8-	.2	4-	2.5	96
0.0	.1-	0	2.2	100

$\Sigma xy = 51.5$ $\qquad$ $2.3 = \sigma x$ $\qquad$ $100 = \mu x$

$0.6 = \sigma y$ $\qquad$ $9.6 = \mu y$

$$\sigma_{xy} = \frac{\Sigma xy}{N\sigma x\ \sigma x} = \frac{51.5}{0.6 \times 9.6 \times 10} = 0.89$$

وهنا، يكون (95%) من العينات التي حجمها (50) فرد مأخوذة من مجتمع حيث ρxy = صفر تقع ضمن المدى (0.0 -2σp , 0.0 + 2 σp) أو بين (-28. إلى 28.): ويجب ملاحظة أن المعادلة (2-15) هي تقريب للخطأ المعياري لمعامل الارتباط والذي يكون دقيقاً فقط لقيم N

الكبيـرة. وقـد اقـتـرح كل من مكنمـار (Mc Nemar, 1962) وهيـز (Hays, 1981) أن هذه الصيغة للخطأ المعياري يجب أن تطبق عندما تكون قيمة N تساوي (50) أو أكثر.

والمحك الثاني لمعنى معامل الارتباط هو قيمة مربع معامل الارتباط ρxy، وهذه القيمة تفسر على أنها نسبة التباين في درجات y المرتبطة بتباين درجات x. فإن كان $\rho xy^2 = .70$، فهذا يعني أن 49% من تباين y مرتبط بالتباين على x ولكن 51% من التباين المتبقي في درجات y غير مرتبط بالفروق الفردية على المتغير x. وطريقة أخرى لتفسير هذا، هو المثال الآتي: افترض أن عالم نفس نمو حسب الارتباط بين عمر الأم ونسبة ذكاء الطفل IQ عند عمر 3 سنوات، وكان الارتباط (.30). والانحراف المعياري لدرجات IQ (=15)، وتباينها (225). ففي هذا المثال 9% من تباين درجات IQ للأطفال مرتبط بعمر الأم. وافترض أن الباحث اختار مجموعة فرعية من هؤلاء الأطفال الذين أعمار أمهم 25 سنة. نتوقع الآن أن تباين درجات IQ لهذه المجموعة الفرعية تكون أقل من (9%) من تباين المجموعة الأصلية إذ لا يوجد تغير مرتبط بعمر الأم، وهنا يكون تباين درجات (IQ) مساوياً (204.75) وانحرافها المعياري (=14.31)، ويعد هذا اختزال قليل للانحراف المعياري لدرجات (IQ) (عند حذف التغير المرتبط بعمر الأم)، وهذا مؤشر إلى أن العلاقة المشار إليها من خلال معامل الارتباط قد تكون ذات فائدة تطبيقية قليلة. لهذا فإننا نرى الارتباط متوسط القيمة يجب تفسيره بحذر إذ أنه مؤشرٌ إلى نسبة مئوية قليلة من التباين المشترك للمتغير قيد الدراسة.

تحذيران آخران يجب ملاحظتهما في تفسير معامل الارتباط الأول، إن اكتشاف علاقة قوية بين المتغيرين (y,x) لا يؤشر وجود علاقة سبب ونتيجة بين المتغيرين. فمثلاً راتب المعلم والمشتريات ومعدل الجريمة في الولايات المتحدة تبين أن هناك ارتباط موجب بين هذه المتغيرات. فهذا لا يعني أن رفع راتب المعلم سيؤدي إلى تصعيد مستويات الجريمة، بينما تفسير أكثر عقلانية هو أن كلا المتغيرين يعكسان عوامل مثل كثافة المجتمع. وبصورة مشابهة الارتباط العالي بين درجات الرضا الوظيفي وراتب العمال لا يمكن تفسيرها لتعني أن مستوى الرضا الوظيفي هو المسؤول عن رواتب العمال. عوامل أخرى متداخلة مثل الاهتمام التنظيمي وأخلاقيات العمال قد يؤشر على كلا المتغيرين. والتحذير الآخر الذي يجب أخذه بعين الاعتبار في تفسير معامل الارتباط هو أن تقييد (أو خفض) التباين على أحد المتغيرين (x أو y) يقلل من أقصى قيمة لمعامل الارتباط يمكن الحصول عليها. مثال على ذلك، افترض أن الطلبة الجدد جميعهم في جامعة كبيرة حصلوا على درجات في الاختبار الفرعي في الرياضيات ترتبط مع درجات اختبار الرياضيات في الفصل الأول وكان معامل الارتباط 0.75. افترض الآن أننا حددنا مجموعة فرعية تحتوي فقط على الطلبة الذين يريدون

التخصص في الرياضيات. من الواضح أن درجات هذه المجموعة على اختبار القبول في الرياضيات أكثر تجانساً (أكثر تقييداً) من درجات المفحوصين جميعهم. كذلك فإن الدرجات المنحرفة على اختبار الرياضيات الفرعي ستكون أقل لهذه المجموعة. فإذا حسبنا معامل الارتباط لهذه الدرجات ذات الانحراف الأقل (مع أن كلا من قيمتي البسط والمقام انخفضتا) نجد أن الارتباط انخفض.

إن أثر تقييد التباين أكبر في المقام، وبالتالي فإن معامل الارتباط سينخفض، وأن معامل الارتباط الملاحظ لهذه المجموعة الفرعية المتجانسة سيكون أقل من 0.75. إن تقييد تباين الدرجات يمكن أن يظهر إذا صمم الفاحص الاختبار وكان صعب جداً أو سهل جداً لأن مثل هذه الاختبارات ستجعل درجات المفحوصين متشابهة. وعند الحصول على معامل ارتباط منخفض ينصح الباحث بالإنتباه إلى إمكانية وجود تقييد في التباين إما بسبب العينة أو بعض مظاهر عملية القياس التي تؤدي إلى علاقة غامضة مضللة ممكنة بين المتغيرات التي نهتم بها، وفي هذه الحالة يفيدنا فحص شكل الانتشار.

التنبؤ بأداء الفرد

عندما يتحقق مستخدم الاختبار من وجود علاقة قوية بين درجات اختبار معين وبعض المتغيرات الأخرى، هنا نرغب باستخدام درجات المفحوصين للتنبؤ بالأداء على متغير آخر. ومثل هذه التنبؤات تعتمد على خط الانحدار والذي يعد أفضل خط مستقيم يمثل شكل الانتشار. ويمكن اشتقاق خطوط الانحدار من الإنتشار المعتمد على الدرجات الزائية أوالدرجات الخام.

خط الانحدار

من معلوماتك الأساسية في الجبر يمكنك تذكر أن معاملة الخط المستقيم تتحدد بالصيغة العامة:

$$y' = bx + c \qquad \text{16-2}$$

حيث تمثل b ثابت يعرف بميل الخط المستقيم و C ثابت يعرف بالقاطع. ويمكن تعريف ميل الخط المستقيم بعدد النقاط على المحور العمودي التي ترتفع (أو تنخفض) لكل زيادة بمقدار وحدة واحدة على المحور السيني. والقاطع هو النقطة التي يلامس بها خط الانحدار المحور الصادي، وهو أيضاً قيمة (y') عند نقطة (س = صفر). وعندما يعبر عن المتغيرات (y,x) بالدرجات الزائية فإنه يمكن كتابة الصيغة العامة لخط الانحدار على النحو الآتي:

$$(Z_{y'}) = m\,(Z_x) \qquad \text{17- 2}$$

بهذه الصلورة تعد المعادلة (2-17) حالة خاصة للصيغة العامة لمعادلة الخط المستقيم عندما يساوي القاطع صفر. وعند فحص انتشار الدرجات الزائية في الشكل (2-5) الذي يبين علاقة خطية يوحي أن كل منهما يتمركز حول الصفر، ولهذا فإن خط الانحدار لأي من انتشارات الدرجات الزائية يجب أن يمر في نقطة الأصل. فإن قيمة القاطع (y) في معادلة انحدار كهذه ستساوي صفر أيضاً. ويوضح الشكل (2-6) خط انحدار مرسوم من خلال انتشار نقاط اعتماداً على الدرجات الزائية.

هذا ويجب على القارئ أن يميز بين الرمزين (Zy\ , Zy)، والتي هي قيم حقيقية لدرجات الفرد الملاحظة. ويتكون خط الانحدار من متصل لمجموعة نقاط، وأعلى من أي قيمة لدرجة الفرد Zx، وقيمة Zy\ هي الدرجة الزائية على المتغير y المُتنبأ بها للفرد عندما يحصل على درجة Zx معينة. لاحظ أنه ولأي قيمة Zx من الممكن الحصول على القيمة المتنبأ لها Zy\ من خط الانحدار، وهذا صحيح حتى لقيم Zx التي لا تظهر في العينة. وملاحظة أخرى هي أنه ولأي قيمة Zx توجد فقط قيمة واحدة لـ Zy\ مع أن انتشار البيانات الفعلية يظهر قيم عدة لـ Zy لدرجة معينة Zx.

وعندما يطابق خط الانحدار الانتشار، فإنه يوجد على الأغلب تناقض بين القيم المتنبأ بها لبعض الأفراد على المتغير y والقيم الفعلية لها. ويسمى هذا التناقض بخطأ التنبؤ وقيمته (Zy - Zy\). ويتبين من الشكل (2-7) ان قيمة Zy\ المتنبأ بها لكارل = 93. ولكن القيمة الفعلية تساوي 1.8. فإذا استخدمنا خط الانحدار للتنبؤ بدرجة كارل فإن خطأ التنبؤ سيساوي 90% من الدرجة الزائية، وإذا استخدمنا خط الانحدار بالدرجات الزائية لكاثي على المتغير y فإن خطأ التنبؤ سيساوي -27. من الدرجة الزائية. من الواضح أن بعض أخطاء التنبؤ ناتج عن استخدام خط الانحدار، فلماذا نهتم بتطوير خط انحدار واحد لمجموعة معينة من البيانات. افترض أن Zx تمثل الدرجة على اختبار القبول في الكلية، و Z_y تمثل معدل الدرجات. فإذا حصلنا على كلا الدرجتين لعينة واحدة من الطلاب (مثل صف عام 1984)، فمن انتشار هذه النقاط يمكن تحديد خط الانحدار الذي يسمح لعمل تنبؤات حول المعدل لمن سيطبق عليه الاختبار مستقبلاً بالاعتماد على درجاتهم على اختبار القبول (Zx). وكخلاصة فعندما تتوافر بيانات عينة حالية على المتغيرين، فإنه يمكن استخدام خط الانحدار للتنبؤ بأداء المفحوصين مستقبلاً والذين تتوافر بياناتهم على أحد المتغيرين فقط.

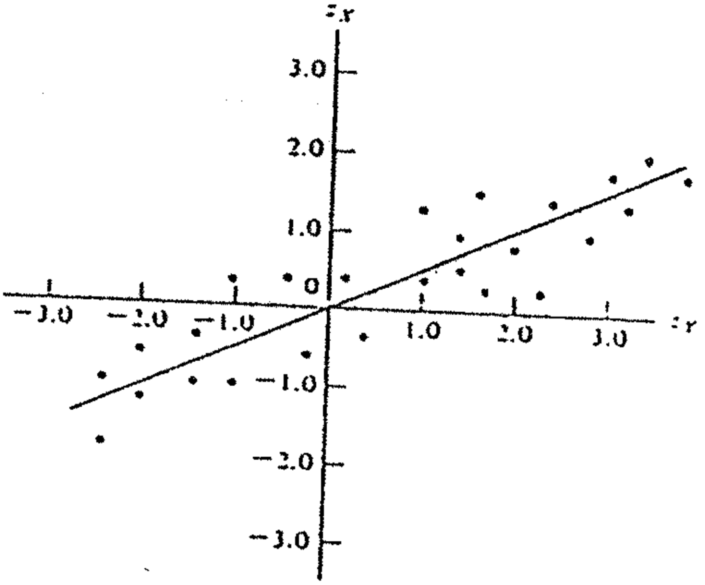
وفي أي محاولة لتمثيل الإنتشار بخط انحدار فإنه لا يمكن تلافي الأخطاء نهائياً. وللحصول على أفضل خط يطابق البيانات فإنه يجب اختيار قيمة m\ (ميل الخط) والتي تقلل هذه الأخطاء، والكمية المختارة لتمثل هذه الأخطاء جميعها هي $\Sum (Zy - Zy\backslash)^2$، وتعرف هذه

القيمة بأنها محك المربعات الأقل. ومع أننا نحاول إيجاد القيمة التي تخفض أقل المربعات إلى أقل قدر بالمحاولة والخطأ ونكون بالوقت نفسه متأكدين بأننا لم نجرِ الاحتمالات الممكنة جميعها. ولحسن الحظ فإنه توجد طريقة رياضية هي إيجاد المشتقة التي تفيد في تحديد القيمة الأدنى للمعادلة الرياضية، والتي عند تطبيقها على المقدار $\sum (Zy - Zy')^2$ تبين أن الكمية تنخفض لخط الإنحدار عندما تكون:

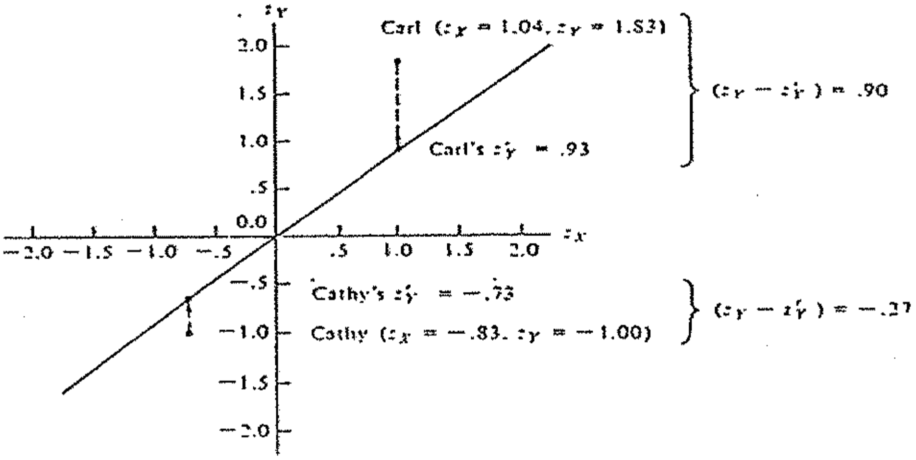
$$\frac{\sum Zx\, Zy}{N} = m$$

والطرف الأيسر لهذه المعادلة هو نفسه في المعادلة (2-12) والذي يعطي صيغة معامل الارتباط، ولهذا يمكننا إعادة كتابة المعادلة (2-17) بتعويض ρxx لـ m لذا فإن:

$$\rho xy\, Zx = Zy' \qquad \text{18-2}$$



شكل (2-6): خط الإنحدار للإنتشار المعتمد على الدرجات الزائية



شكل (2-7): الفجوات بين قيم Zy الفعلية و \Zy المتنبأ بها لمفحوصين اثنين.

معادلة الإنحدار للدرجات الخام:

كما أن لمعامل الارتباط صيغة رياضية لحسابها باستخدام الدرجات الخام، فإنه يوجد كذلك صيغة للتنبؤ بالدرجات الخام \Y من قيم درجات x المعروفة. ولبناء العلاقة بين الدرجة الزائية والدرجة الخام لمعادلة التنبؤ دعنا نمثل \Zy على أنها:

$$Zy' = \frac{y' - \mu y}{\sigma y}$$

ومن المعادلة (2-18) نعرف أن $Zy' = \rho xy \, Z_x$ ، وبتعويض هذه القيمة نصل إلى أن:

$$\frac{y' - \mu y}{\sigma y} = \rho xy \left(\frac{x - \mu x}{\sigma x} \right)$$

وبالمعالجة لقيمة \y نحصل على:

$$\rho xy \, (x - \mu x) \, \sigma y / \sigma x = y' - \mu y$$

$$\mu y + \rho xy \, (x - \mu x) \, \sigma y / \sigma x = y'$$

ومع أن هذه المعادلة تظهر معقدة، إلا أنه يمكن إجراء تبسيطين عليها، في الأول دعنا نحدد:

$$\rho xy \quad \sigma y / \sigma x = by.x \qquad 19\text{-}2$$

ويطلق على الثابت by.x معامل الإنحدار، وبالتالي يمكننا إعادة كتابة معادلة الدرجة الخام على النحو الآتي:

$$Pyx\ (x - \mu x) + \mu\ y = Y^{\backslash}$$

وبإجراء عملية الضرب نحصل على:

$$by.x\ (x) - by.x\ (\mu x) + \mu y = Y^{1} \qquad 20\ -2$$

والآن، دعنا نحدد ثانية القيمة C على أنها:

$$-by.\ x\ (\mu x) + \mu y = C$$

ومنها نحصل على:

$$\rho xy\ (x) + c = Y^{\backslash} \qquad 21\ -2$$

وهذه مطابقة للمعادلة العامة للخط المستقيم.

ولنوضح ما سبق الآن بحساب معامل الإنحدار والقاطع لخط الإنحدار للتنبؤ بالمعدل من درجات اختبار الاستعداد باستخدام القيم المبنية في الجدول (4.2)، يمكننا حساب:

$$.0.06 = (\frac{0.6}{9.6}) \quad (0.89) = by.x$$

والقاطع:

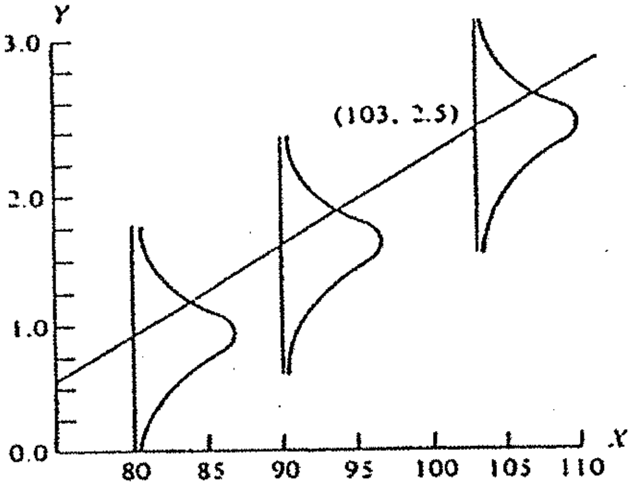
$$3.7- = 2.3 + (100)\ .06- = C$$

ومن السهل باستخدام معادلة الانحدار هذه حساب الدرجة المتنبأ بها $Y^{\backslash}$ لأي مفحوص عند معرفة درجته (x). لذلك للمفحوص ذو درجة x = 1.3 فإن:

$$(3.7-) + (6.3)\ (.06) = Y^{\backslash}$$

$$2.5 =$$

ويبين الشكل (2-8) خط الانحدار لدرجات (y,x) الخام للبيانات المدونة في الجدول (2-4).



شكل (2-8): توزيع درجات y الملاحظة حول قيم مُعطاة لـ Z و $Z^{\backslash}$.

لاحظ أنه عندما نستخدم الدرجات الخام لا يمكن أن تكون قيمة القاطع C (صفراً)، لذا فإن خط الإنحدار لا يمر بنقطة الأصل، كذلك فإن ميل خط الإنحدار للدرجات الخام ($\rho xy\ \sigma y / \sigma x$) من الممكن أن يكون أكبر من (1.00). وأخيراً، وللتبسيط يمكن أن نبدأ من قيم (y,x) على المحورين عند نقطة غير الصفر.

الخطأ المعياري للتقدير:

عند استخدام معادلة الإنحدار للتنبؤ بدرجات الأفراد فمن المناسب أن نهتم بمقدار الخطأ لهذا التنبؤ. ويوضح الشكل (2-8) توزيع قيم (x) الفعلية حول قيم ($y^{\backslash}$) لعدة قيم مختلفة على المتغير x. ومثل هذا التوزيع يطلق عليه أسم التوزيع الشرطي، وذلك لأنه يمثل توزيع قيم (y) لقيمة (x) معينة. وكل من هذه التوزيعات له إنحراف معياري، وعلى افتراض تساوي الإنحرافات المعيارية الشرطية فإن كل منهما يعرف بـ الخطأ المعياري للتقدير. وعند التنبؤ بـ ($y^{\backslash}$) من قيمة معينة لـ (x) فإن الخطأ المعياري للتقدير يرمز له بـ $\sigma_{y.x}$ ويحسب بـ:

$$\sigma_{y.x} = \sigma_x \sqrt{1 - \sigma_{xy}} \qquad (22\text{-}2)$$

ولبيانات جدول (4.2) فإن قيمة $_{y.x}$ تكون:

$$\sigma_{y.x} = 0.6 \sqrt{1 - 0.89} = 0.27$$

ولاستخدام الخطأ المعياري للتقدير في تفسير درجات الاختبار نفترض أن:

1) تتوزع أخطاء التنبؤ طبيعياً حول كل قيمة $y^{\backslash}$.

2) توزيع أخطاء التنبؤ متشابهة لقيم x جميعها.

ولهذا يمكننا الإفتراض أنه إذا كانت $\sigma_{y.x} = .27$ فإن 68% من القيم المتنبأ بها لـ $y^{\backslash}$ التي تساوي 2.5 ستقع درجاتهم ضمن الفترة ($2.5 \pm 1\sigma$) أي من (2.23 إلى 2.77)، وتقريباً فإن 95% منها ستكون درجاتهم ضمن إنحرافين معياريين أي بالفترة من (1.96 إلى 3.04). ومثل هذه التقديرات تساعدنا في رؤية التنبؤ المشكوك بصحته لدرجات الأفراد على أساس معادلة الانحدار.

الخلاصة:

تم في هذا الفصل مراجعة الجداول التكرارية والمنحنيات على أنها إحدى طرائق تلخيص توزيع الدرجات، بالإضافة إلى الإحصاء الوصفي الذي يمكن استخدامه في تلخيص المظاهر الأساسية لتوزيع الدرجات. وعلى الأقل يستخدم نوعين من الإحصاءات: مقاييس النزعة المركزية ومقاييس التشتت. وهنالك ثلاثة مقاييس شائعة للنزعة المركزية هي المتوسط والوسيط والمنوال. ويوصف تشتت البيانات عادةً بوساطة المدى والتباين والإنحراف المعياري. وقد تم عرض صيغ حساب كل من المتوسط والإنحراف المعياري للمجتمعات المحدودة وغير المحدودة، وكذلك توضيحاً لاستخداماتها.

فعندما يتحدد موقع فرد على توزيع معين تستخدم الدرجة الزائية لتؤشر إلى عدد الإنحرافات المعيارية التي تقع عندها درجة المفحوص لأعلى أو أدنى من متوسط المجموعة. كذلك تم عرض صيغة دالة المنحنى الطبيعي، وتم تعريف التوزيع الطبيعي المعياري، وقيم التوزيع الطبيعي المعياري بالدرجات الزائية واحتمالية ظهور أي درجة أقل أو تساوي قيمة Z التي يمكن الحصول عليها من جدول الدرجات الزائية الطبيعي المعياري. وتعد هذه الطريقة الأكثر ملائمة في تفسير درجة الاختبار لأن أي توزيع تكراري مقارب للمنحنى الطبيعي تكون نسبة المفحوصين أعلى أو أدنى من درجة معينة يمكن الحصول عليها من الجدول الزائي الطبيعي المعياري.

وعندما نهتم بالعلاقة بين متغيرين فإن معامل الارتباط هو الإحصائي الوصفي المناسب. وتتراوح قيمته بين (-1.00 إلى +1.00). وتؤشر قيمة هذا العدد إلى

الدرجة	التكرار للمجموعة 1	التكرار للمجموعة 2	التكرار للمجموعة 3
28	1	3	1
27	2	7	4
26	3	5	3
25	7	1	4
24	5	1	3
23	1	2	4
22	1	1	1

2/ احسب المتوسط، التباين، الإنحراف المعياري لكل من التوزيعات الثلاث.

3/ أكمل جدول القيم لكل مفحوص في المجموعة الأولى.

الإسم	الدرجة الخام	الدرجة المنحرفة	الدرجة الزائية
جون	22	-	-
بيتر	__	3	-
ادوارد	__	-	1.45
كاتي	24	-	-

4/ لكل من التوزيعات الثلاث، ما قيم الدرجات الخام المناظرة للدرجات الزائية صفر، وما قيم الدرجات الخام المناظرة للدرجات الزائية – 1.00؟

5/ اعتماداً على إجابتك للسؤال الرابع: هل من المنطق القول أنه لدرجة زائية معينة تكون الدرجة الخام المكافئة في المجموعة الثالثة دائماً أقل من الدرجة الخام للمجموعة الأولى؟ فسر استدلالك؟

6/ افترض أن المفحوصين الأربعة اختيروا كجزء من مجموعة كبيرة يقترب توزيعها من

مدى إمكانية تمثيل العلاقة بين المتغيرين الأول والثاني بخط مستقيم وبفعالية، وإشارة الارتباط باستخدام الدرجات الخام أو الدرجات الزائية تؤشر إلى نوع العلاقة. وفي تفسير الارتباط فإن مطوّري الاختبارات ومستخدميها قد يجدوا مربع الارتباط مفيداً كمؤشر لكمية التباين المشترك بين المتغيرين. ومن المهم أيضاً اختبار الدلالة الإحصائية للارتباط المحدد من العينة.

وحالما تتحدد درجة العلاقة بين المتغيرين فمن المرغوب به استخدام درجة المفحوص على المتغير الأول في التنبؤ بالأداء على المتغير الآخر. ومثل هذه التنبؤات تعتمد على استخدام خط الإنحدار. وخط الانحدار لمجموعة البيانات هو أفضل خط ينطبق عليه التمثيل بالأزواج المرتبة للنقاط على أزواج الدرجات على المتغيرين. وعندما يعتمد التمثيل بالأزواج المرتبة على الدرجات الزائية فإن ميل خط الانحدار يكون معامل الارتباط فقط. وفي حالة اعتماده على الدرجات الملاحظة فإن معامل الإنحدار للمتنبيء (y من x) يرمز له بالرمز $b_{y.x}$، وهو نتاج لمعامل الارتباط و σ_x / σ_y وقيمة القاطع لخط الإنحدار للدرجات الزائية = صفر، وقيمة القاطع لخط الإنحدار للدرجات الخام يكون $(-b_{y.x}\,\mu_x + \mu_y)$ وعند استخدام صيغة معادلة الخط المستقيم لتحديد قيم (y) المتنبأ بها لمفحوص درجته (x) معروفة تظهر بعض أنواع الخطأ في التنبؤ. والخطأ المعياري للتقدير هو مقياس للتشتت الذي يظهر في القيم الملاحظة لـ (y) حول القيمة المتنبأ بها لـ ($y\backslash$) للمفحوصين الذين لهم درجة (x) نفسها.

التمارين:

1/ افترض أنه تم تطبيق اختبار أدائي على ثلاثة مجموعات من المفحوصين في مواقف مختلفة. وتم تدوين التوزيعات التكرارية على النحو المبين أدناه. أجب عن الأسئلة الآتية دون إجراء أية حسابات:

أ) بصورة عامة، أي المجموعات يبدو أن أدائها كان الأفضل في الاختبار.

ب) أي المجموعات يظهر أنها الأكثر تجانساً في المجموعات الثلاث.

جـ) لأي المجموعات يبدو التوزيع التكراري أكثر شبهاً بالتوزيع الطبيعي.

د) ولأي المجموعات يبدو الاختلاف الأكبر بين المتوسط والوسيط للدرجات.

التوزيع الطبيعي. والدرجات الزائية لبعض أفراد المجموعة على النحو الآتي:

دييبيكا	2.15-
شارون	2.15
رونالد	1.50
شيلدون	.70

أ) تقريباً، ما نسبة المفحوصين الذين درجاتهم أقل من درجة دييبيكا؟

ب) تقريباً، ما نسبة المفحوصين الذين درجاتهم أعلى من شارون؟

جـ) تقريباً، ما نسبة المفحوصين الذين درجاتهم بين رونالد وشارون؟.

د) أي من درجات للمفحوصين الأربعة تظهر بدرجة أكبر في توزيع المجموعة الكلية؟

7/ طلب من طلبة الدراسات العليا تحويل الدرجات الخام للمجموعة الثالثة في التمرين الأول إلى درجات زائية تناظرها، واستخدم أحد التلاميذ الصيغة:

$$Z = \frac{x - \mu}{\sigma}$$

في حين استخدم تلميذ آخر التكرار التراكمي لكل درجة وقسمها على N، ونظر بعدها إلى قيمة Z المناظرة في جدول Z– الطبيعي المعياري. فهل يحصل كلا التلميذين على النتائج نفسها؟ ولماذا؟

8/ أ– افترض أن المتغير x يتوزع طبيعياً وبمتوسط قدره 0.50 وانحراف معياري 10، فما قيمة:

$$\int_{45}^{55} f(x)\,dx$$

حيث تمثل f(x) معادلة المنحنى الطبيعي.

ب– أعد صياغة المعادلة في 8– أ كما يتم تطبيقها على توزيع درجات الاختبار (للمساعدة: أنظر تمرين 6 جـ).

9/ باستخدام الجدول الآتي والمتضمن درجات أربعة اختبارات منفصلة لسبعة مفحوصين حولت درجاتهم جميعاً إلى الدرجات الزائية.

المفحوص	اختبار الاستعداد	الاستيعاب القرائي	الوعي الاجتماعي	تقدير الذات
الأول	.00	.42-	.58-	1.44-
الثاني	.49	.42	1.36	.39-
الثالث	.41-	.00	.39	.79
الرابع	1.06	1.13	.19	1.05
الخامس	.81-	.99-	1.17	.52-
السادس	1.38	1.41	1.17-	1.58
السابع	1.71-	1.55-	1.36-	.52

وضح بالتمثيل البياني العلاقة بين الاستعداد والاستيعاب القرائي، وبين الاستعداد والوعي الاجتماعي، وبين الاستعداد وتقدير الذات. واعتماداً على هذه المنحنيات أي زوج من المتغيرات تتوافر فيه أقوى علاقة خطية، وهل هي موجبة أم سالبة؟

10/ ما هي معاملات الارتباط بين أزواج المتغيرات الثلاث في التمرين التاسع؟ وما نسبة التباين في الاستيعاب القرائي المشترك مع التباين في الاستعداد؟ وما نسبة التباين المشترك بين تقدير الذات والاستعداد؟

11/ بمعرفة الارتباط بين الاستعداد وتقدير الذات المبين في تمرين 10، اقترح استاذ أن تحسين الوعي الاجتماعي يرفع من درجات اختبار الاستعداد، فهل توافقه الرأي ولماذا؟ افترض أن الاستاذ اقترح تحسين الاستيعاب القرائي ويجب أن يرفع درجات الاستعداد، الآن هل توافقه الرأي؟

12/ ما درجة Z المتنبأ بها في الاستيعاب القرائي لمفحوص درجته على اختبار الاستعداد Z = -.75؟ وما درجة Z المتنبأ بها لهذا المفحوص في تقدير الذات، وإن كان متوسط الدرجات الخام في اختبار تقدير الذات 20 وانحرافها المعياري 7.62، فما الدرجة الخام المتنبأ بها؟

13/ أجرى باحثان مختلفان دراسات على العلاقة بين دافعية التحصيل وموقع الضبط الداخلي. ودوّن أحد الباحثين الارتباط الذي حصل عليه أنه أكبر وبدلالة احصائية وبقيمة 35. من الصفر عند درجة ثقة 95% . ودون الباحث الآخر ارتباط قيمته 0.35 وقال أنه ليس أكبر وبدلالة إحصائية من الصفر وعند درجة ثقة 95% فكيف يمكن تفسير هذا التناقض؟ وهل وجد الباحثون درجة العلاقة الخطية نفسها بين هذين المتغيرين؟

14/ طبق مقوِّم برنامج اختبار قبلي وآخر بعدي على مسترشدين اشتركوا جميعاً في البرنامج الإرشادي وحصل على البيانات الآتية:

	الاختبار القبلي	الاختبار البعدي	قبلي – بعدي
المتوسط	20	36	$.63 = \rho$
الانحراف المعياري	5	3	

فإذا حصلت جانيت على درجة 15 في الاختبار القبلي، فما القيمة المتنبأ بها لدرجتها على الاختبار البعدي؟ وإذا حصل توم على درجة 15 في الاختبار القبلي فهل يكون له الدرجة البعدية المتنبأ بها نفسها مثل جانيت؟ وهل لهذين المفحوصين الدرجة البعدية الفعلية نفسها؟ وهل يمكنك إعطاء مدى لفرق الدرجات المتوقع بين درجتيهما البعدية؟

15/ إذا حصل مفحوصان على درجات تبعد نقطة (درجة) عن بعضها في الاختبار القبلي. فكم تبعد درجاتهم المتنبأ بها على الاختبار البعدي؟ وإذا حصل مفحوصان على درجات تبعد 5 نقاط عن بعضها في الاختبار القبلي، فكم تبعد عن بعضها في الدرجات المتنبأ بها على الاختبار البعدي.

16/ إذا أردت أن ترسم خط انحدار للدرجات الخام في التمرين الرابع عشر: فما القيم التي ستستخدمها لتمثيل الميل والقاطع لهذا الخط؟ وإذا أردت رسم خط الانحدار لدرجات Z فما القيم التي ستستخدمها لتمثل الميل والقاطع.

17/ افترض أن مشرفاً أعطى اختباراً قصيراً مؤلفاً من (15) فقرة وحسب نقطة واحدة لكل منها. وكان المتوسط (10) درجات والانحراف المعياري (3) وكان معامل ارتباط درجات الاختبار مع درجات الاختبار القصير السابق (0.56). افترض أن المشرف قرر فيما بعد أن يعطي لكل فقرة وزناً يساوي (5) نقاط لذا فإنه ضرب درجة كل مفحوص بالعدد 5.

أ– ما أثر هذا على المتوسط.

ب– ما أثر هذا على الدرجة المنحرفة للمفحوص.

جـ– ما أثر هذه الإضافة على التباين.

د– كيف تؤثر هذه الإضافة على معامل الارتباط مع الاختبار السابق.

18/ افترض أن درجات توزيع ما ضربت جميعها بقيمة ثابتة k، أي أن:

$$kx = x'$$

$$k\mu x = \mu_{x'}$$

بين أن:

$$k^2\sigma x = \sigma_{x'}^2$$

الفصل الثالث

3

مقدمة الى التدريج

الفصل الثالث

مقدمة الى التدريج

عند مراجعتنا لتعريفات القياس سنلاحظ أن إعطاء أرقام لخصائص الأشياء يجب أن يخضع لقوانين معينة. وتعرف عملية تطوير قوانين منظمة ووحدات ذات معنى لقياس كميات الملاحظات التجريبية بالتدريج. وقد تم اختراع تدريج لقياس الصفات الفيزيائية في المجتمعات البدائية. ومن الضروري تطوير وحدات قياس مناسبة ومقبولة لتكميم الخصائص الملاحظة للأشياء في مجالات التجارة والزراعة والبناء. فعلى سبيل المثال استخدم قدماء المصريين وحدة الذراع لقياس الطول، والتي حددت على أنها طول الساعد من المرفق إلى قمة الإصبع الأوسط، ويمكن عد عدد الوحدات المكونة لشيء ما. وبهذا فإنه يتكون التدريج عندما تتحدد مجموعة القيم الممكنة التي يمكن استخدامها أثناء عملية القياس، وتأسيس قانون واضح لهذه القيم. ويجب أن يتضمن قانون التعيين هذا وحدة القياس المستخدمة في التدريج. ويظهر الاستخدام المزدوج لمصطلح (Scale) بمعنى تدريج أو مقياس في بعض مجالات القياس التربوي والنفسي. ويطلق هذا المصطلح على أدوات قياس معينة مثل مقياس وكسلر أو مقياس القلق الرياضي، وتستخدم هذه الأدوات لجمع عينة منتظمة من السلوك، وأطلق عليها في الفصل الأول اسم اختبارات. وهذا لا يتعارض مع استخدام مصطلح التدريج- المقياس- الذي يظهر في هذا الفصل.

الأعداد الحقيقية وتداريج القياس

تذكر أن الأعداد الحقيقية تتألف من الصفر والأعداد الممكنة والأعداد العشرية جميعها ما بين سالب وموجب وإلى ما لا نهاية. ويمكن تمثيل نظام الأعداد الحقيقية بيانياً ليمثل خط الأعداد الحقيقية، وهو متصل واحد يمكن تجزئته إلى أجزاء أصغر وأصغر لا نهائية. وكل قيمة في هذا النظام لها موقع واحد ومميز على هذا المتصل.

ولخط الأعداد الحقيقية خصائص مهمة عديدة، سنذكرها على النحو الآتي:

الأول: يوجد مركز ثابت للقيم ويؤشر الصفر لموقع هذا المركز. وعندما تكون مجموعة الأرقام لها نقطة صفر ثابتة نقول بأن المجموعة لها خاصية المركز الثابت. ومن المهم ملاحظة

أن العدد صفر يسلك بطريقة مختلفة بعض الشيء عن أي قيمة أخرى في نظام الأعداد الحقيقية. فعند إضافة الصفر إلى أي عدد آخر ينتج العدد الآخر نفسه، وعلى هذا فإن القيمة الأصلية لا تتغير بإضافة الصفر لها مع أن قيمتها تتغير بإضافة أي رقم آخر لها.

الثاني: الوحدة الأساسية لنظام الأعداد الحقيقية هي (1) (عنصر التماثل). وبالإنتقال بعيداً عن المركز فإن الأعداد المتتابعةعلى خط الأعداد الحقيقية تقع على بُعد وحدة واحدة من العدد الذي يسبقها. وحجم هذه الوحدة ثابت (ويمكن تصوّرها على أنها المسافة بين صفر وواحد)، لذا فإن المسافة بين أي عددين متتابعين على خط الأعداد وأي نقطة تكون متساوية.

الثالث: عند مقارنة أي قيمتين على خط الأعداد الحقيقية فإن القيمة المطلقة للعدد الأبعد عن الصفر يُعد الأكبر لأنه يمثل تراكم عدد أكبر من الوحدات من العدد الذي يكون أقرب إلى صفر. وعلاوة على ذلك فإن العلاقة بين القيم في نظام الأعداد الحقيقية هي علاقة تعدي. أي أنه ولثلاثة قيم مختلفة (أ، ب، جـ)، إذا كانت أ > ب، و ب > جـ يجب أن نستنتج أن أ > جـ. وعندما تكون هذه العلاقة صحيحة لمجموعة أعداد نقول أن هذه المجموعة لها خاصية الترتيب. وعندما تكون هذه من خصائص نظام الأعداد فإنه من الممكن تحديد العمليات الحسابية جميعها (الجمع والطرح والضرب والقسمة) ويمكن اشتقاق قوانين العمليات التجميعية والترابطية والتوزيعية لها.

ويعمل قانون التدريج كرابطة أو قانون للمطابقة بين العناصر في نظام البيانات والعناصر في نظام الأعداد الحقيقية. ويمكن تحديد النظام على أنه تجمع عناصر أو أشياء تشترك بخاصية معينة (Weitzenhoffer, 1951). وسنستخدم مصطلح نظام البيانات ليشير إلى مجموعة الملاحظات الممكنة لخاصية معينة لمجموعة أشياء. افترض أن لدينا مجموعة أشياء مختلفة في أطوالها، ويمكننا تكوين قواعد تدريج يمكن استخدامها في الربط بين الأطوال الملاحظة في نظام البيانات مع القيم على خط الأعداد الحقيقة. فمثلاً يمكن ترتيب الأشياء من الأقصر إلى الأطول والعدد الذي يشير إلى أي شيء يعتمد على موقعه، فالشيء الأقصر يُعطى الرقم 1، والذي يليه في القصر يعطي الرقم 2، وهكذا ... ويمكن أن تقارن الأشياء بوحدة معيارية، مثل وحدة القدم الموجودة على المسطرة وإعطاء اعداد تعتمد على الأطوال بوحدة القدم. وعندما يتم تحديد قانون التدريج واعطاء أعداد لكل عنصر من بيانات النظام يطلق على هذه الأعداد اسم قيم التدريج. وقد تكون الخصائص الأساسية أو العلاقات التي تظهر في نظام الأعداد الحقيقية ذات فائدة وقد لا تكون.

مستويات تدريج القياس

حدد ستيفنز (Stevens, 1946) أربعة مستويات لتدريج القياس تختلف فيما بينها بمدى

احتفاظ قيم التدريج بخصائص خط الأعداد الحقيقية. وهي: الإسمي والرتبي والفئوي أو الفاصلي والنسبي أو المطلق.

التدريج الأسمي وتستخدم الأعداد فيه لتصنيف العناصر في نظام البيانات، ولا يكون لها خصائص الترتيب، أو المسافات المتساوية بين الوحدات أو المركز الثابت. فالأعداد على قمصان لاعبي كرة القدم أو على التذاكر تُعد أمثلة عليها. ويستخدم علماء الاجتماع أحياناً التدريج الإسمي في ترميز الاستجابات في بنود الاستبانة الديموغرافية. فمثلاً يرمز لبند جنس المستجيب المذكر بالعدد صفر وللأنثى بالعدد 1، مع أن لا يوجد تمييز كمي بين الجنسين في التطبيق. وتشير هذه الأعداد إلى اختلافات وصفية للفئات الممثلة للإستجابات وتخدم كرموز أو تصنيفات مختصرة لهذه الأشياء. ومع التدريج الإسمي يمكن استخدام أية مجموعة من الأرقام ولكن كل شيء أو استجابة في بيانات النظام له عدد مختلف عن غيره. ويمكن استبدال مجموعة الأعداد بأخرى شريطة الاحتفاظ بالمطابقة بين مجموعتي الأعداد، ويطلق على هذا اسم التحويل المتماثل.

افترض على سبيل المثال أن الاستجابات على فقرة تتعلق بالمذهب الديني أعطت الأعداد على النحو الآتي: 1 للبروتستانت و2 للكاثوليك و 3 لليهود و 4 لأي ديانة أخرى. يمكن إجراء تحويل متماثل واستخدام الاعداد الآتية: 8 للبروتستانت، و6 للكاثوليك و 9 لليهود و1 لأي ديانة أخرى، ولا يسمح في عملية التحويل إعطاء الرقم 2 للبروتستانت والكاثوليك إذ أن لكل منهم رمز خاص به في قانون التدريج الأصلي. أيضاً يسمح التدريج الإسمي بحرية كبيرة في عملية تحويل مجموعة قيم التدريج إلى أخرى دون خسارة أية معلومات عن العلاقة بين البيانات الأصلية، وهذا ممكنٌ لأن قيم التدريج الأصلي التي تؤشر للبيانات (فاقدة لـ : الترتيب والمسافة والمركز) تحوي القليل من المعلومات الكمية للبدء بها.

وعندما يتم ترتيب العناصر في بيانات النظام وفق كمية الخاصية المقاسة، وتطلب قانون التدريج قيماً من نظام الاعداد الحقيقية التي تؤشر بالترتيب نفسه ينتج التدريج الرتبي، الرتبي، وأحد الأمثلة الشائعة للتدريج الرتبي في العلوم الطبيعية هو تدريج رموز لصلابة المعادن والذي يستخدمه علماء المعادن في وصف درجة صلابة المعادن. وهذا التدريج مستخدم منذ عام 1822، ويعطي العدد 1 للكلس و 2 للجبس و 3 للكالسيت وهكذا حتى 10 للماس. وأُعطي كل معدن القيمة العددية اعتماداً على قدرته على خدش سطح أي معدن له رقم أقل على تدريج الصلابة. فالمعدن الذي يخدش الجبس ولا يخدش الكالسيت مثلاً يُعطى القيمة 2.5. وكما هو ملاحظ في أحد أدلة الجيولوجيا فإن هذا التدريج لا يمثل علاقة رياضية تامة. فالقيمة 10 على التدريج تعني فقط صلابة أكثر من 9 ، و 9 أكثر صلابة من 8 وهكذا (Pough, 1960,p.31). ومثال آخر، الحكم على ملكات الجمال إذ ترتب الكونتيسات من 1

إلى 10 باستخدام معايير الشخصية. وتتضمن هذه القيم تدريج رتبي. لاحظ وجود صعوبة في التمييز بين الكونتيسات الأولى والثانية وصعوبة أقل نسبياً في تقدير الرتب الخامسة والسادسة وهذه المعلومات لا تغطيها قيم هذا التدريج. وعلى هذا فإن قيم التدريج الرتبي لها خاصية ترتيب القيم نفسه في نظام الأعداد الحقيقية، ولكنه يفتقر إلى خاصيتي المسافات المتساوية بين الوحدات والمركز الثابت. ويمكن تحويل القيم على التدريج الرتبي إلى قيم أخرى باستخدام أي قانون شريطة الاحتفاظ بالمعلومات الأساسية عن رتب عناصر البيانات. ويطلق على مثل هذا التحويل رتيب أو مماثل. فإذا رتبت خمسة عناصر تقارن بين بيانات النظام أساساً (1 , 2 , 3 , 4 , 5) فإنه يمكن استبدالها بالأعداد (3 , 10 , 35 , 47 , 201)، من جهة أخرى فإن التحويل إلى (10 , 16 , 12 , 18 , 20) على التوالي للعناصر نفسها لا يجوز إذ أنه تم تغيير ترتيب العنصر الثاني والثالث بالنسبة إلى مواقعها الأصلية.

وتشير الأعداد في التدريج الفئوي أو الفاصلي إلى الرتب إضافة إلى أن المسافات بين الأعداد لها معنى بالنسبة للخاصية المقاسة، فإذا وجدت درجتين عند النهاية الدنيا للمتصل وكان الفرق بينهما وحدة واحدة، ووجدت درجتين عند النهاية العظمى للمتصل وكان الفرق بينهما وحدة واحدة، فإن الفرق بين الدرجتين المنخفضتين يمثل كمية الخاصية نفسها في الفرق بين الدرجتين عند النهاية العظمى. لاحظ أن هذه العلاقة ليست صحيحة في تقدير الكونتيسات على التدريج الرتبي. أيضاً للتدريج الفئوي خصائص الترتيب والمسافات المتساوية بين الوحدات، وتبقى نقطة الأصل لمثل هذا التدريج اصطلاحية ولا تمثل الغياب الكلي للخاصية المقاسة. للتوضيح تخيل أربعة من المسجونين في خلية ولديهم طاولة وورق اللعب فقط، وطرح سؤال عن أطوال هؤلاء ولكن ليس لديهم مسطرة. ولتحديد من هو الأطول يجب أن يقيس كل منهم الاخر ووحدة القياس هي طول ورقة اللعب، ولأن ليس لديهم عدد كافي من الورق لقياس الطول بدءاً من الأرض فقد اختاروا ارتفاع الطاولة كنقطة بداية، وتم قياس طول كل منهم بعدد من أطوال أوراق اللعب بدءاً من ارتفاع الطاولة. وتم التعبير عن أطوالهم المقاسة بطول 12 بطاقة، وطول 11 بطاقة، وطول 10.5 بطاقة، وطول 13 بطاقة. لاحظ أن طول أي شيء يمكن اختياره كنقطة بداية، وهنا يتوافر وصف الأطوال على تدريج فئوي، فهذه القياسات منسوبة لمركز مناسب (ومع هذا فإن هؤلاء يعرفون أطوالهم المطلقة).

ويعد تدريج ميزان الحرارة الفهرنهيتي مثال لمقياس تدريجه فاصلي. فقد يندهش القارىء لأن هذا التدريج له نقطة صفر، لكن نقطة الصفر هذه ليست مثل صفر نظام الأعداد الحقيقي. وهو أن كمية الطاقة الحرارية الممثلة بالصفر الفهرنهيتي لو أضيفت إلى كمية أخرى من الطاقة الحرارية فإن الكمية المتوافرة من الحرارة ستزداد. بكلمات أخرى لا تمثل درجة الصفر على ميزان الحرارة الفهرنهيتي الغياب الكلي للحرارة وبالتالي فإن الصفر هنا لا يسلك مثل الصفر في نظام الأعداد الحقيقة. ولأن القيم النسبية للمسافات التي تؤشر قيمة الملاحظات

(القياسات) تتضمن معلومات لها معنى، فإن التحويلات المسموح بها لقيم التدريج الفاصلي تكون مقيدة. فإن كانت x تمثل قيم على التدريج الأصلي و y تمثل قيم على التدريج المُحوَّل فإن التحويل الوحيد الذي يحتفظ بالمعلومات المتضمنة في مجموعة القيم الأصلية هو ص = أس + ب حيث أ و ب قيم ثابتة. لذا فإن القيم (5 , 4 , 2 , 3) يمكن تحويلها لتصبح (11 , 9 , 5 , 7) وذلك بإجراء التحويل ص = 2 س + 1، ودون فقدان أية معلومات تتعلق بالبناء الذي تجري دراسته.وإذا حاولنا إجراء تحويل غير خطي مثل ص = س2، فإن القيم 25 , 16 , 4 , 9 لا تكون العلاقة بينها هي العلاقة نفسها بين (5 , 4 , 2 , 3) على التدريج الأصلي.

المستوى الرابع للقياس هو التدريج النسبي أو المطلق، وهذا له خصائص الترتيب والمسافات المتساوية بين الوحدات والمركز الثابت أو نقطة الصفر المطلق والعديد من مقاييس الأشياء الفيزيائية هي مقاييس نسبة (مثل الطول بوحدة السنتيمتر، والوزن بالباوند، والعمر بالأيام والشهور والسنين). وأطلق على التدريج اسم النسبي لأن موقع الصفر المطلق معروف. ويمكن التعبير عن القياسات غير الصفرية على هذا التدريج كنسبة احداها إلى الأخرى. أحد الأمثلة المشهورة على هذا التدريج هو ميزان حرارة كلفن، إذ تتحدد نقطة الصفر في عند الإنعدام الكلي للحرارة. والتحويل الوحيد على التدريج النسبي ويحفظ المعلومات من التبديل هو ص = جـ س حيث جـ مقدار ثابت. ويجب ملاحظة أن أي مجموعة من القيم تحقق متطلبات التدريج النسبي يمكن اعتبارها تحقق متطلبات التدريج الفاصلي (وذلك إذا أردنا فقدان المعلومات المتعلقة بموقع الصفر المطلق). وبصورة مشابهة فإن البيانات المجتمعة على التدريج الرتبي، والبيانات المجمعة على تدريجات رتبية أو فاصلة أو نسبية فإنها بالتأكيد تحقق متطلبات التدريج الأسمي.

ومن المهم تذكر أن مستويات تدريج ستيفنز للقياس تؤلف تصنيف مفيد لكنه ليس الطريقة الوحيدة لتصنيف أو وصف المستويات. فعلى سبيل المثال تحدث تورغرسن (Torgerson, 1958, P. 19) عن مستويين للتدريج الرتبي، وقدم كومبس (Coombs, 1950b) نوع من التدريج يقع بين الرتبي والفاصلي. وتطرح دائماً قضية للمناقشة حول القياس في الاختبارات النفسية عما إذا كان من المستوى الرتبي أو الفاصلي. وسنناقش هذه القضية لاحقاً في هذا الفصل ببعض التفصيل. وأولاً فمن الضروري فهم عمليات التدريج التي تطبق على البيانات النفسية بشكل أوسع.

أساليب التدريج في تطوير القياس:

عند تطوير أداة للتقييم التربوي أو النفسي فإن مطور الاختبار يعمل على اختبار سلسلة من الفرضيات حول تدريج البيانات التي حصل عليها من قياسه للبناء المقترح.

الأول: يجب صياغة الفرضية بوضوح، وهي أن البناء له خاصية تظهر بكميات مختلفة، لذا فإنه يمكن تكميمها باستخدام قانون التدريج المقترح على متصل نظري أحادي الإتجاه، وعادةً يطلق عليه اسم متصل نفسي.

الثاني: السؤال عن ماهية خصائص الأعداد الحقيقية (الترتيب، المسافة، المركز) لقيم التدريج على المتصل. وميز تورغرسون (1958) بين ثلاث طرائق واسعة يمكن للباحث أن يستخدمها في فحص هذه القضايا: طرائق تتمركز على الأفراد، وطرائق تتمركز على المثير، وطرائق تتمركز على الإستجابة.

طرائق تتمركز حول الفرد:

هنا يكون اهتمام مطور الاختبار الأساسي وضع الأفراد عند نقاط مختلفة على المتصل. ويعد هذا هدف أكثر قياسات الاستعداد والتحصيل، وهو الهدف الأساسي في بناء العديد من الأدوات في المجال الإنفعالي. فمثلاً افترض أن محلل نفسي طور مقياس للإحباط، يمكن لهذا المحلل أن يكتب 20 فقرة وتصنف الاستجابات على أنها تؤشر على مشاعر الإحباط أولاً. وبعد ذلك يؤشر المستجيب وتعطى له الدرجة 1 للإستجابة التي تؤشر للإحباط والدرجة صفر للاستجابة التي لا تؤشر، وتحسب درجة كل فرد من جمع درجات الفقرات جميعها. فالأفراد الذين لهم درجات عالية يتحدد موقعهم أقرب إلى النهاية الموجبة للمتصل أكثر ما هو للأفراد الذين درجاتهم أقل. ويتحدد وزن درجات الفقرات بقرار مناسب من مطور الاختبار، وعلى الأغلب تعطى الفقرات أوزان متساوية. وقد يعطي القليل من الإنتباه للفروق المحتملة بين الفقرات التي سيجيب عليها الفرد وذلك لأن الهدف الأساسي هو تدريج الأفراد. وحتى عندما يسمح للأفراد الإجابة عن كل فقرة مع بعض التدرج في مستوى الاستجابة (مثل صيغة ليكرت للاستجابات إذ يختار المفحوص استجابة تتراوح بين موافق بشدة إلى غير موافق بشدة). وحتى هذا يمكن أن يمثل طرق تتمركز حول الأفراد. وكما لاحظ تورغرسون (1958) أن لهذه الطريقة فوائد تطبيقية عدة إلا أنها لا تؤدي إلى تطوير نماذج تدريج تسمح للباحث باختبار خصائص تدريج الدرجات المشتقة. وتحسب الدرجة الكلية بناء على افتراض بسيط، وهو أن القيم التي تؤشر لكل صيغة استجابة ممكنة تشكل تدريج عددي له خصائص رتيبة أو رتيبة بوحدات متساوية.

ويبدو أن اختيار طريقة نفسية تتمركز حول الفرد تسمح للباحث بأن يحصل عليها بسهولة. وعند هذه النقطة لا يوجد اختبار مقنع للخصائص العددية للبيانات المستخلصة، وسيواجه الباحث بهذا السؤال دائماً. فالدرجات المأخوذة من المقاييس النفسية يجب اختبارها

في الحال للتحقق من ثباتها وصدقها (ويوضح الفصل الرابع نظرة عامة لعملية تطوير الاختبار والتحقق من صدق وثبات المقاييس التي تتمركز حول الفرد، وتطرح الوحدات الثانية والثالثة هذه الطرائق أيضاً). وعندما لا تكون الدرجات المستخلصة بهذه الطريقة من التدريج ليست رتبيّة تماماً، ولها وحدات قياس متساوية تقريباً فإنها لا تحقق معايير صدق وثبات مقبولة. وعند هذه النقطة فإن مطور الاختبار لا يعرف سبب فشل الأداة حول ما إذا كان بسبب قصور خصائص الأعداد في التدريج أو عدم ملائمة الصدق والثبات في القياس للأهداف التي خضعت للاختبار.

طرائق تتركز حول المثير

هنا يهتم الباحث في تحديد موقع المثيرات أو (الفقرات) على المتصل النفسي. وكانت التطبيقات المبكرة لهذه الطرائق في المختبرات التجريبية لعلماء الإدراك الألمان في عام (1800). وأطلق على هؤلاء الباحثين اسم السيكوفيزيائيين لأن اهتمامهم كانت في تطوير علاقات كمية بين الاستجابة على المثيرات الفيزيائية والمثيرات نفسها. ومن هؤلاء السيكوفيزيائيين أرنست ويبر وغوستاف فخنر الذين تركز اهتمامهم في تحديد الفروق الصغيرة في المثيرات الفيزيائية (الضوء، نغمة الصوت، الوزن،...الخ) والتي يمكن اكتشافها بثبات بالحواس البشرية. فعلى سبيل المثال لقياس إدراك الضوء، فمن المألوف تقديم إضاءتين وسؤال الأفراد أيهما أكثر إضاءة. ومع أن المجرب يمكنه تحديد ذلك بقياس فيزيائي للمثير الأقوى، إلا أن المشكلة تكمن في تطوير طريقة لتدريج استجابات الأفراد، وهو: كيف يمكن تدريج متصل نفسي لشدة الإضاءة؟مع أنه تم تطوير طرائق عدة، فإن احداهما وهي الأفضل وتتم بمقارنة إضاءتين مرات عديدة من قبل الفرد نفسه أو بوساطة مجموعة أفراد، ويمكن اعتبار المثيرين متكافئين على تدريج الحواس وذلك عندما يتم اختيار كل مثير بالتكرار نفسه (50% من المرات)، وعندما يتم اختيار مثير معين لأكثر من 75% من المرات فإن الفرق بين هذين المثيرين على المقياس نفسه يطلق عليه «الفرق الملاحظ Just Noticeable difference) (JND). واختير محك 75% لأنه يقع في منتصف المسافة بين الصدفة والدقة التامة. لقد تم استخدام «الفرق الملاحظ» (JND) على أنه وحدة قياس من قبل علماء النفس الأوائل على المتصل النفسي وباستخدام طريقة التدريج العامة هذه ومع تنقيحات عديدة لها قدم ويبر القانون الخاص به (قانون ويبر) وحاول اثباته تجريبياً، وهذا القانون ينص على أن الزيادات المتساوية في قوة المثير ينتج عنها نسب زيادات مساوية لها في القيمة في الاستجابة الحسية للفرد. واقترح فخنر بعدها تعديل باقتراحه لعلاقة لوغاريتمية بين زيادة الاحساس والزيادة في المثير الفيزيائي. ومع أن السيكوفيزيائيين بذلوا الكثير من الجهد في دراسة تقنيات

تجريبية مختلفة لتحديد وحدة «الفرق الملاحظ»، ونادراً ما تعرضوا للسؤال عما إذا كانت هذه الوحدة ذات وحدات متساوية على المتصل النفسي. وقد لاحظ تورغرسون (1958) أن وحدة «الفرق الملاحظ» ذات وحدات متساوية نفسياً من خلال التعريف.

وبقيت محاولات تدريج المقاييس النفسية محصورة ضمن علماء النفس الإدراكي حتى ظهور جهود ثيرستون الرائدة إذ أنه أعد طرائق تدريج تناسب مقاييس الإتجاهات. فقد بين ثيرستون أنه من الممكن تدريج خصائص المثير التي لا ترتبط بأي مقياس فيزيائي. فعلى سبيل المثال: فقد بين أنه يمكن تدريج الجدية في الجرائم وذلك بسؤال القضاة (الخبراء) أن يختبروا أزواج الجرائم الممكنة في القائمة جميعها (مثل حرق المباني والتهريب، حرق المباني والقتل العمد،...الخ) وتحديد أي الجرائم أكثر جدية في مقارنة كل زوج منها (Thurstone,1927). دراسة أخرى لثيرستون كانت عرض مجموعة من الفقرات المنظمة المتعلقة بالمذهب الديني وطلب من الأفراد تحديد موقعها على متصل مقسم إلى أجزاء متساوية بالاعتماد على قوة إنفعالاته (مشاعره) تجاه الكنيسة، والمعبر عنها في الفقرات (Thurstone& Schave, 1929). وقد جذب عملية تطبيق التدريج على المثيرات اللفظية اهتمام كل من علماء الاجتماع والتربية وعلم النفس. واقترح ثيرستون تقنية كمية مصاحبة لعملية اختبار ما إذا كانت قيم التدريج المشتقة من هذه الطرائق تتطابق مع نموذج التدريج الفاصلي. وإحدى هذه الطرائق المستخدمة لدمج اختبار ما تسمى «قانون الأحكام المقارنة»، وطريقة أخرى تسمى «قانون تصنيف الأحكام». وتعد قوانين ثيرستون بالفعل أنظمة تستخدم قوانين في تقدير قيم التدريج لكل مثير. وخصص تورغرسون (1958) عدة فصول لكل طريقة من طرائق التدريج المتمركزة حول المثيرات بما تحتويه من طرق جمع بيانات ومعادلات تستخدم في تقدير قيم التدريج والاختبارات الإحصائية لحسن المطابقة بين قيم التدريج الملاحظة والمقدرة، ومثل هذه التفصيلات ليست من أهداف هذا الفصل. ومع ذلك سنصف باختصار كيف يمكن تطبيق قانون الأحكام المقارنة عندما تكون البيانات المجمعة هي أحكام تستخدم مقارنات الأزواج المحتملة جميعها بين الفقرات.

أولاً: يجب فهم بعض الشيء عن النظرية التي تتضمنها هذه الطريقة. افترض أن متصل نظري يهتم في تقدير المسافة بين فقرتي اتجاه نحو استعمال الأسلحة النووية. والفقرة الأولى (1) قد تكون مصاغة على النحو الآتي:

«يجب استخدام السلاح النووي كحل أخير في الحرب»

يفترض ثيرستون أن الأفراد المختلفين لهم آراء متباينة تتعلق بقوة هذه الفقرة، وتتوزع هذه

الآراء طبيعياً وتتمركز حول نقطة على متصل الإتجاه يرمز لها بالرمز (μ i) (وهذا التوزيع لا يمكن ملاحظته مباشرة). والآن لنفترض أن الفقرة (ل) هي:

«لا يجب استخدام السلاح النووي تحت أي ظروف ممكنة».

يفترض أيضاً أن آراء الأفراد حول قوة هذه الفقرة تتباين وبتوزيع طبيعي يتمركز حول النقطة (μ_J). إن موقع كل من μi و μ_J تعد ممثلة للموقع الحقيقي لكلا الجملتين على متصل الإتجاهات. أخيراً لاحظ أنه ولكل فرد يمكن التعبير عن المسافة بين موقع الفقرتين i و j بـ (j-i). وهذه المسافات تتباين أيضاً بتوزيع طبيعي، ويطلق على هذا التوزيع اسم توزيع الفروقات المميزة. ومتوسط هذا التوزيع هو:

$$\mu j - \mu i = \mu d$$

والانحراف المعياري لهذا التوزيع هو σ_{j-i}. ويخدم هذا الانحراف المعياري كوحدة قياس في تحديد المسافة بين i و j على متصل الاتجاهات من خلال التعبير:

$$\frac{\mu j - \mu i}{\sigma_{i-J}} = Zij$$

وهذا التعبير هو في الحقيقة الدرجة الزائية والتي تعبر عن الفرق المعياري بين μi و μj وهو أيضاً المسافة المعيارية بين المثيرات i و j على المتصل النفسي.

ويكون هدف الباحث في تدريج فقرات مقياس الاتجاهات أكثر من مجرد زوج واحد من الفقرات. وعلى هذا فإن المسافات المعيارية تحسب لجميع أزواج الفقرات، فمثلاً لمثيرات ثلاثة نحصل على:

$$\frac{\mu_2 - \mu_1}{\sigma_{12}} = Z_{12}$$

$$\frac{\mu_3 - \mu_1}{\sigma_{13}} = Z_{13}$$

$$\frac{\mu_3 - \mu_2}{\sigma_{23}} = Z_{23}$$

وأشرنا مسبقاً إلى أن التدريج الفاصلي له نقطة أصل مناسبة. ولحساب قيم التدريج في المثال الحالي على تدريج مناسب فمن الضروري تحديد نقطة أصل مناسبة. دعنا نعتبر أن μi

= صفر، لذا فإن قيمة الفقرة 1 على التدريج = صفر (ولا يمكن اعتبار هذا التدريج بشدة تدريج نسبي). والآن يمكننا القول بأن:

$$Z_{12} = \frac{\mu_2 - \text{صفر}}{\sigma_{12}}$$

$$Z_{13} = \frac{\mu_3 - \text{صفر}}{\sigma_{13}}$$

$$Z_{23} = \frac{\mu_3 - \mu_2}{\sigma_{23}}$$

وبإعادة ترتيب هذه المعادلات نحصل على:

$$\mu_2 = \sigma_{12}\ (-Z_{12})$$

$$\mu_3 = \sigma_{13}\ (-Z_{13})$$

$$\mu_3 - \mu_2 = \sigma_{23}\ (-Z_{23})$$

وإذا عرفنا قيم Z_{12} ، و Z_{13} ، و σ_{12}، و σ_{13}، فإنه من الممكن حل المعادلتين الأولى والثانية والحصول على قيم التدريج للفقرات (3,2). وكما نرى فإنه من الممكن حساب كل من Z_{12}، Z_{13}، Z_{23} وليس σ_{12} و σ_{13}. وقد تم حل هذه المشكلة باستخدام إحدى صيغ قانون مقارنة الأحكام، وذلك بافتراض أن الانحرافات المعيارية الأثنين متساوية. وبالرمز للإنحراف المعياري بالرمز k تصبح معادلاتنا:

$$\text{صفر} = \frac{\mu_1}{12}$$

$$-Z_{12} = \frac{\mu_2}{K}$$

$$-Z_{13} = \frac{\mu_3}{K}$$

ومن الضروري تبين أن هذا ممكناً لأننا افترضنا أن μ_1 = صفر، وبالتالي فإننا على الأكثر نحصل على ترتيب فئوي. ونحن نعرف أن التدريج إذا كان فئوياً فمن الممكن تحويل كل قيمة على التدريج إلى أخرى وفقاً للصيغة $a\,(\mu i) + b$. وبالتعويض صفر لقيمة (b) و $1/k$

لقيمة a يمكننا تحديد مجموعة قيم للتدريج الجديد μ_i ترتبط يمة بقيم μ_i^* الأصلية بالمعادلة $\hat{\mu}_i = 1/_k (\mu_i)$ + صفر.

μ_i^*= صفر

$\mu_{12}^* = Z_{12} -$

$\mu_{13}^* = Z_{13} -$

وتعد قيم μi المحوّلة لكل فقرة مؤشراً لموقعها على المتصل بالنسبة للفقرة التي اعتبرت نقطة أصل مناسبة. وفي مثالنا هذا تبقى مشكلة واحدة هي حساب قيم Z12 و Z13 وذلك للحصول على قيم μ_2^* و μ_3^* على التوالي.

وعلى القارىء ملاحظة أن كل ما نعرفه عن طبيعة التدريج والمعبر عنه بقيم μ_i^* ليس تدريج نسبي وإنما فئوي. وتزودنا طريقة ثيرستون للأحكام المقارنة بأساس للحصول على قيم Zij من الملاحظات التجريبية، واختبار ما إذا كانت هذه الحسابات تتبع التدريج الفئوي. فعند تحديد موقع مثيرين i و j على النقطة نفسها على المتصل وسؤال مجموعة خبراء للتأشير على أيهما أقوى، فإننا نتوقع أن يشير نمط الإستجابات إلى بعض التشويش في تقدير الأحكام. خصوصاً أننا نتوقع أن 0.50 من الأحكام تختار الإستجابة i > j و 50. تختار الاستجابة j>i. وبزيادة المسافة بين (μ j-μ i) يقل التشويش ويصبح من المحتمل أن 0.75 من الأحكام تختار i>j وفقط 25. منها تختار j>i. وبزيادة المسافة أكثر تجد أن توزيع الأحكام للإستجابات يصبح Pij = 0.95 و Pji = 05.، حيث أن Pij تمثل نسبة الأحكام التي تختار i>j و Pji تمثل نسبة الأحكام التي تختار j>i. وبالتالي فإن هنالك علاقة بين احتمالية اختيار الحكم i > j عشوائياً (Pij) والمسافة (μ j-μ i). وبتطبيق قانون ثيرستون للأحكام المقارنة فإنه يفترض أن يكون توزيع الاستجابات عبر الأحكام في هذا الموقف ثنائي الاختيار محكوماً بالتوزيع الطبيعي للفروق المميزة. وهذا يعني أن قيم Pij يمكن أن ترتبط بالدرجات الزائية باستخدام جدول الاحتمالية المعياري الاعتدالي... لذا على سبيل المثال نظرياً، عندما تكون:

Pij = .50 فإن Zij = 0.00

و Pij = .75 فإن Zij = 0.68

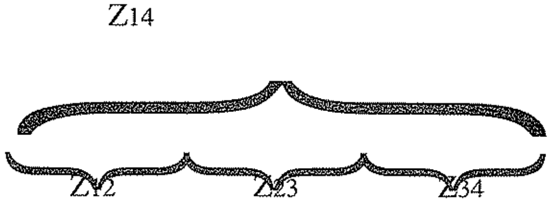
و Pij = .95 فإن Zij = 1.64 ... وهكذا.

وعندما يكون لدينا فقرتان فقط فإننا لا نذهب إلى أبعد من هذا وغالباً يكون لدينا في المقاييس أكثر من فقرتين، وهنا نسأل المحكمين المقارنة بين الأزواج الممكنة جميعها، وإجراء كل تحكيم باستقلالية عن الآخر. وبهذا يكون لدينا مجموعة أولية لتقديرات Zij لكل زوج من

الفقرات. دعنا نرى الآن كيف يمكننا استخدام هذه البيانات في حساب قيم μ_i لكل فقرة، واختبار ما إذا كانت هذه القيم تتطبق على تدريج فاصلي.

1. افترض أن الشكل (3-1) يبين المواقع الحقيقية لأربع فقرات على متصل نفسي، وافترض أنه تدريج فاصلي. وبافتراض أن قيم σ_{ij} جميعها متساوية، فإن قيم Z_{ij} تساوي المسافات بين أي مثيرين (J,i). وبالتالي فإن ما يأتي يكون صحيحاً:

$$Z_{34} + Z_{23} + Z_{12} = Z_{14}$$



شكل (3-1):

العلاقة الجمعية لقيم Z_{ij} حيث أن قيم أزواج ij للفقرات جميعها تم تدريجها على متصل فئوي

وفي الواقع، حتى عندما تكون قيم μ_i تقع على تدريج فئوي فإن الأحكام على أزواج المثيرات المختلفة تكون مستقلة.ونادراً* ما تتطابق قيم Z_{ij} الأولية التي نحصل عليها من مقارنات الأحكام للأزواج المحتملة كلها وذلك بسبب الأخطاء العشوائية لعملية التحكيم.

2. إن قانون الأحكام المقارنة يزودنا بمجموعة معادلات تستخدم قيم Z_{ij} الأولية في حساب قيم Z_{ij}، والتي تتفق مع شروط التجميع بأقل انحراف عن القيم الأصلية لـ Z_{ij}. والافتراض هو أن التقديرات المنقحة تمثل إلى درجة أقرب الصورة الحقيقية لمواقع الفقرات النسبي، والذي يظهر في حالة التخلص من الأخطاء العشوائية لعملية التحكيم (ويوجد لهذا القانون صيغ عدة تعتمد على الافتراضات المتعلقة بتوزيعات الأحكام).

3. إن الفجوة بين قيم Z_{ij} التجريبية وتلك المقدرة يمكن اختبارها باستخدام الاختبارات الإحصائية لحسن* المطابقة. فإن كانت الفروق قليلة فإن المتصل النفسي يستخدم لتقدير مطابقة المثيرات للتدريج الفئوي. والذي نفترضه عند استخلاص قيم Z_{ij}. وأما إذا كانت الفروق كبيرة جداً (وذلك بمحك حسن المطابقة الإحصائي) فيجب أن نتساءل حول افتراض قوة التدريج على أنه فئوي للمتصل المستخدم في تدريج قيم Z_{ij}. وكما لاحظ تورغرستون (1958) فإن الهدف من الاختبار الإحصائي هو تحديد ما إذا كان التدريج الناتج تام أو أنه بالتقريب يصل إلى درجة يمكن معها معالجة البيانات على

متصل الأعداد الفئوي دون الإبتعاد عن ذلك كثيراً.

ويمكن استخدام المجموعة الأخيرة من الفقرات المدرجة الآن بهدف القياس. وغالباً فإن الفقرات تقدم لمجموعة مستجيبين بصيغة موافق- غير موافق، وتحسب درجات كل مستجيب على أنها متوسط درجات الفقرات التي أجاب عنها. ويجب ملاحظة أنه تم اختيار قانون الأحكام المقارنة هنا لأنه يزودنا بتوضيح بسيط لكيفية اختبار مقارنة البيانات التجريبية لنموذج تدريج متمركز حول المثير. وفي تاريخ تطوير تدريج مقاييس الاتجاهات فقد حل محل هذه الطريقة قانون تصنيف الأحكام لأن هذه الأخيرة لا تتطلب من الأفراد استهلاك وقت كثير في تقدير الأحكام لأزواج الفقرات المحتملة كلها، ولكنها تتطلب باختصار ترتيب أو تصنيف أو وضع فقرات المثير في فئات.

طرائق تتمركز حول الاستجابة

وهي الطريقة الأصعب في التدريج، إذ أنها تستخدم بيانات الاستجابة في تدريج الأفراد على المتصل النفسي بالاعتماد على قوة الفقرات التي أجاب عليها (أو أجاب عنها بصورة صحيحة). وفي الوقت نفسه يتم تدريج الفقرات من خلال قوة السمة أو كميتها عند الأفراد الذين يجيبون على هذه الفقرات.

وسنقدم توضيحين لهذه الفئة من طرائق التدريج. فقد وصف جوتمان (Guttman, 1941b, 1950) طريقة تدريج الاستجابة المعروفة باسم تحليل التدريج. وفي الغالب تستخدم هذه الطريقة عدد قليل غير متحيز من الفقرات. وقد تم ترتيب هذه الفقرات حسب الزيادة في القوة، لذلك فإن وافق مستجيب على إحدى الفقرات فيجب عليه أن يوافق على الفقرات الأخرى جميعها التي تعبر عن إتجاه (أو شعور) إيجابي أضعف نحو البناء وأن لا يوافق على الفقرات التي تظهر إتجاه سلبي (معاكس). وتوضيح لمثل هذه المجموعة ما يأتي:

أ- يحق لرجل الدين في الخدمة العامة أن يضرب.

ب- يحق للمعلم في المدرسة العامة أن يضرب.

جـ- يحق للمرضات في مستشفيات الولاية والمدنية أن تضرب.

د- يحق لرجل الإطفاء أن يضرب.

تسمى الاستجابات المتسقة منطقياً نماذج الاستجابات المسموح بها. ويبين جدول 3-1 النماذج المسموح بها للمثال السابق. وتسمى نماذج الاستجابة غير المتسقة أخطاء. وكلما كان عدد الأفراد الذين تتطابق استجاباتهم أكبر كلما كان التأكيد أكثر على أن هذه الفقرات تكوّن

تدريج رتبي. ويقاس مدى ذلك بمعامل الإنتاج C.

$$C = 1 - \frac{\text{عدد الاخطاء الكلي}}{\text{عدد الاستجابة الكلي}}$$

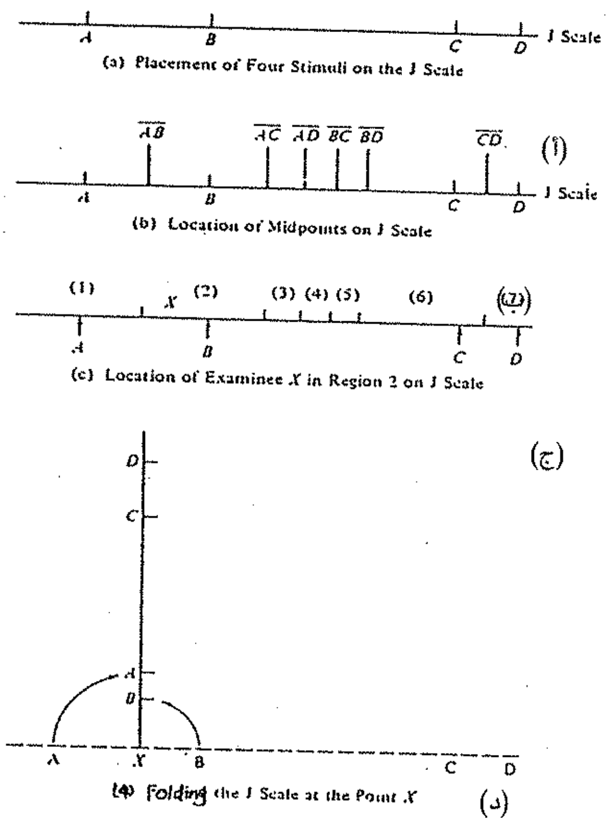
ويستخدم القيمة 0.90 كمحك لإثبات التدريج الرتبي للفقرات (Torgerson, 1958, p. 323). وطرق تقدير المسافة بين الفقرات على هذا المتصل تكون أكثر تعقيداً ولن نبحثها في هذا الفصل، وسنبحث في الفصل 15 النماذج اللوغاريتمية لتدريج الاستجابات على الفقرات، وللتدرب على تطبيق المفاهيم الأساسية لتدريج جوتمان فإنها طرحت في تمارين نهاية الفصل.

جدول (3-1): أنماط الاستجابة المسموح بها على تدريج جوتمان لفقرات الضرب الأربعة من قبل المستخدمين في الدوائر العامة.

الانماط المسموح بها	الجمل			
	A	B	C	D
1	+	+	+	+
2	+	+	+	−
3	+	+	−	−
4	+	−	−	−
5	−	−	−	−

اقترح كومبس (Coombs, 1950b) طريقة أخرى لتدريج كل من الأفراد والفقرات تسمى تقنية الانتشار. وهنا نطلب من المستجيب أن يقدر ترتيب تفضيلاته لمجموعة مثيرات أو أن يقدر ترتيب مجموعة فقرات حسب مدى تقاربها مع معتقداته الشخصية «ولفهم النظرية المرتبطة بهذه الطريقة دعنا نفترض إن لدينا مجموعة مؤلفة من أربعة مثيرات مدرجة، وأن موقعها الحقيقي على المتصل النفسي معروف لنا (وذلك من خلال الإظهار)، ويسمى تدريج هذه المثيرات بالتدريج الرابط أو تدريج J، إذ أننا نستطيع تحديد موقع كل من المثير والأفراد عليه. ويرمز لمواقع المثيرات الأربعة على تدريج J بالحروف من (A إلى D) في الشكل (3-2، أ). ويبين الشكل (3-2، ب) النقطة المتوسطة بين أزواج المثيرات، ففي هذا الشكل تشير النقطة AB إلى النقطة التي تتوسط المثيرين A و B، والنقطة AC تتوسط المسافة بين المثيرين A و C وهكذا.

ويوجد ستة أزواج من المثيرات المتجمعة وبالتالي توجد ستة نقاط تتوسط المسافات على التدريج. وهذه النقاط الستة تقسم التدريج J إلى سبعة مناطق. ونظرياً إن كان موقع الفرد في المنطقة (1) عندما يُسأل عن تقدير الترتيب للمثيرات الأربعة بالنسبة لموقعه الخاص



شكل (3-2): توضيح لموقع المثير بطريقة الإنتشار لكومبس

فإنه سيعطى الاستجابات D C B A. وهذا يكون أكثر وضوحاً إذا تخيلنا المفحوص x في المنطقة 2 (أنظر شكل 3-2 جـ). وإذا قمنا بثني المتصل عند موقع المفحوص x فإنها ستظهر كما في شكل (3-2د). وبذا يمكن ترتيب تفضيلات المفحوص x للمثيرات الأربعة، فيمكن قراءتها مباشرة من الثني العمودي للمتصل كما تبين في شكل (3-2د). ويسمى ترتيب المثيرات على المتصل العمودي تدريج I (الفرد). وفي حالة توافر مثيرات مدرجة بالضبط

وأفراد مدرجين بالضبط فإنه سيكون لدينا سبعة نماذج مسموح بها لـ الفرد -الاستجابة (أو تدريجات I)، واحدة لكل منطقة على تدريج J. وفي مثال تدريجات I السبعة الممكنة فإنها تكون على النحو الآتي:

موقع الفرد (تدريج J)	انماط الاستجابة (تدريج I)
المنطقة 1	DCBA
المنطقة 2	DCAB
المنطقة 3	DACB
المنطقة 4	ADCB
المنطقة 5	ADBC
المنطقة 6	ABDC
المنطقة 7	ABCD

وفي تطوير التدريج النفسي لا تتوافر لدينا معرفة عن موقع المثيرات على تدريج J. وبدلاً من ذلك فإننا نبدأ بمجموعة تتكون من أربعة مثيرات لا نعرف إن كان بالإمكان ترتيبها وما يجب أن يكون عليه الترتيب. فعلينا محاولة عمل هذه الاستدلالات من خلال نماذج الفرد – الاستجابة الملاحظة. ولتوضيح ذلك دعنا نعود إلى فقرات الإتجاه الأربعة حول الضرب من قبل المستخدمين في المؤسسات العامة. فنحن سنفترض أن كل منها كتبت على بطاقة منفصلة، ورمزت كما يأتي:

رجال الدين. رجل الدين في المؤسسات العامة له الحق في أن يضرب .

المعلمين. المعلمين في المدارس العامة لهم الحق في أن يضربوا.

المرضات. المرضات في مستشفيات المدنية والولاية لهم الحق في أن يضربوا.

رجال الإطفاء. رجال الإطفاء لهم الحق في أن يضربوا.

افترض الآن أننا طبقنا العبارات الأربعة على عينة مؤلفة من 200 طالب كلية، وطلبنا منهم تقدير ترتيب البطاقات وذلك بوضع الجملة التي تمثل شعورهم بدرجة أقوى في هذه القضية أولاً ثم الجملة الثانية وهكذا ..... فإن كانت هذه العبارات قابلة للتدريج، فإننا سنحصل من

خلال نموذجنا على سبعة نماذج استجابة (تدريج I)، وقد يشير الحصول على أكثر من سبعة نماذج إلى غياب قابلية التدريج للمثيرات أو الأفراد أو كليهما (ومع الأفراد الواقعيين فإن عدد قليل من نماذج الإنحراف أو الأخطاء يمكن أن تظهر). ويبين الجدول (3-2) تكرارات الإستجابة التي تم الحصول عليها. وبالنظر المتمعن فإن البيانات تشير إلى أنه تم اختيار سبعة من التدريجات المكنة لـ I تم اختيارها بتكرار عالٍ. ومنها جميعاً فإن اثنتين فقط تعد إحداهما مرآة للأخرى:

رجال دين- معلمين- ممرضات- رجال إطفاء ورجال إطفاء- ممرضات- معلمين رجال دين. وتسمح هذه النتيجة باستدلال حول ترتيب المثيرات الأربعة على المتصل، ويمكننا الآن ترميز العبارات بالحروف A B C D لتشير إلى ترتيب أماكنها على التدريج J.

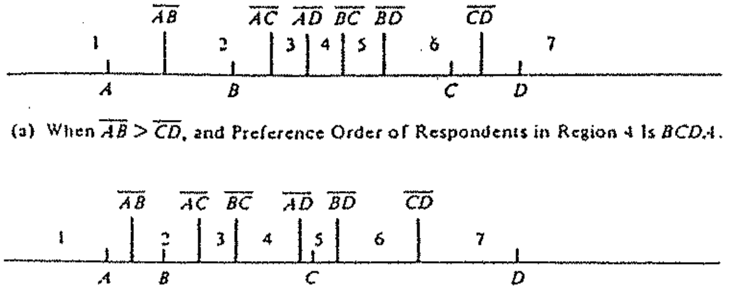
جدول (3-2)" التكرارات التوضيحية لأنماط الإستجابة المختلفة التي تم الحصول عليها من تدريج كومبس لأربعة فقرات تقيس الإتجاه نحو الضرب من قبل مستخدمي المؤسسات العامة.

أنماط الاستجابة	التكرار	اعادة ترميز انماط الاستجابة بعد تحديد ترتيبها
رجال دين - معلمين - ممرضات- رجال إطفاء	41	A B C D
معلمين- ممرضات - رجال دين- رجال إطفاء	32	B C A D
ممرضات- معلمين- رجال اطفاء- رجال دين	30	C B D A
ممرضات- رجال اطفاء- معلمين- رجال دين	29	C D B A
معلمين- رجال دين - ممرضات- رجال إطفاء	24	B A C D
رجال إطفاء- ممرضات - معلمين- رجال دين	20	D C B A
معلمين- ممرضات- رجال اطفاء- رجال دين	18	B C D A
معلمين- رجال اطفاء- رجال دين- ممرضات	3	B D A C
ممرضات- رجال اطفاء- رجال دين- معلمين	2	C D A B
معلمين- رجال دين- رجال اطفاء- ممرضات	1	B A D C

(والقرار أي منها سيكون نقطة النهاية لنرمز لها بالرمز A بشكل مناسب، وهنا سيخصص A B C D لتمثل رجال الدين- معلمين- ممرضات- رجال اطفاء، وهكذا فإن A

ترمز إلى أقل تفضيل نحو الضرب و D إلى أعلى تفضيل للضرب، وفحص آخر لتدريج هذه المثيرات لاحظ أن أول تفضيلين في كل من التفضيلات السبعة هي نقاط متتالية على التدريج.

وعندما يتحدد ترتيب المثيرات الأربع فإننا نود استخلاص بعض الاستنتاجات المتعلقة بالمسافات النسبية بينها. وتحديداً يجب أن نجيب على السؤال الآتي وذلك بما يتعلق بالمثيرات الأربع: أي أزواج المثيرات أقرب لبعضها البعض هل هي A و B أم C و D؟ وفي مثالنا هذا نعتقد أنها تكون: رجال الدين والمعلمين أو الممرضات ورجال الاطفاء بلغة حقوقهم في الضرب؟. أولاً: تذكر أنه يوجد سبعة مناطق للأفراد على المتصل. ويجب علينا أن نميز بأن مواقعهم تتحدد بالمسافات بين أزواج المثيرات. وعندها يوجد أربع مثيرات فقط فإن موقع المنطقة 4 (المنطقة المتوسطة) يعتمد على ما إذا كانت المسافة بين A و B أكبر من المسافة بين C و D أو العكس. ولنرى هذا قارن بين الأشكال (3-3أ) و (3-3ب)، فالفرق بين كلا الطرفين ينعكس على الترتيب المقدم من قبل الأفراد في هذا التصنيف. وعلى هذا فعندما تؤدي النتائج التجريبية إلى نموذج إستجابة معين (تدريج I) مثل B C D A فإننا نستنتج أن المسافة النفسية بين A و B تتجاوز المسافة بين C و D. وهذا هو الحال للبيانات الموضحة في جدول (2-3) والممثلة بالشكل (3-3أ). بالمقابل إذا كانت المسافة النفسية بين C و D أكبر من المسافة بين A و B فإن شكل 3-3 ب يمثل هذه الحالة، ونتوقع بالتالي نموذج استجابة C B A D.



(a) When $\overline{AB} > \overline{CD}$, and Preference Order of Respondents in Region 4 Is *BCDA*.

(b) When $\overline{AB} < \overline{CD}$, and Preference Order of Respondents in Region 4 Is *CBAD*

شكل (3-3): مواقع المناطق المتوسطة على تدريج J عندما نعكس المسافات بين BA و CD من حيث قيمتها.

وهذا هو الاستنتاج الوحيد الذي يمكن تكوينه عن المسافة بين المثيرات في حالة أربع مثيرات على تدريج J. وفي حالة خمس مثيرات على تدريج J يمكننا تكوين 12 مجموعة بأحد عشر نموذج استجابة لكل منها. وباتباع الطريقة المنطقية نفسها مع المثيرات الأربعة فإن لدينا المعلومات الكافية لتحديد المسافات النسبية عبر المثيرات جميعها. وبين كومبس (Coombs, 1950b) أن هذا النوع من التداريج من المحتمل أنه يقع بين التدريجين الرتبي والفاصلي، وذلك

لأن المسافات بين الوحدات لم تحدد بدقة، ولكن الحجم النسبي لهذه المسافات هو المعروف. وفي هذا المثال المحدود استخدمنا الفحص البصري والمنطقي لاختبار مطابقة البيانات للنموذج، ومع ذلك توجد أساليب تحليل معقدة لتحديد مدى مطابقة الاستجابات للنموذج المثالي أنظر: (Torgerson, 1958, Ch.14).

والقاريء المهتم بالتطبيقات المتخصصة لتقنيات التدريج في بناء مقاييس الاتجاهات يمكنه الاستعانة بمصدر مثل (Dawes, 1972) و (Udinsky, Osterlind, Slynch, 1981). ولهؤلاء المهتمين بالتسلسل التاريخي لتدريج الاتجاهات يمكنهم الإفادة من (Fish bein, 1967). ولتغطية بتفاصيل أكثر لنماذج التدريج تجدها في (Torgerson, 1958) و (Van der Ven, 1980).

مستويات التدريج للقياسات المتمركزة حول الفرد

كما رأينا فإنه عندما يبنى تدريج ويختبر بهدف تدريج المثيرات أو تدريج الأفراد والمثيرات فإنه يوجد طرائق متعددة مناسبة لاختبار المدى التي يمكن من خلالها اعتبار القياسات رتبية أو فاصلية. وعند استخدام الطريقة المتمركزة حول الفرد فإنه لا توجد طريقة مناسبة لاختبار ما إذا كانت القياسات ستعالج على أنها معلومات من مستوى رتبي أو فئوي.

ويقدم أحياناً دليل لدعم الإنطباع بأن بيانات الاختبار ذات تدريج فئوي هو العلاقة بين التدريج والتوزيع الطبيعي، ويأخذ هذا الدليل إحدى صيغتين. الأولى هو أنه عند قياس سمات فيزيائية على مقاييس فئوية تكون القيم الناتجة للعينات الكبيرة متطابقة مع المنحنى الطبيعي، ويتم الحصول عادةً على مثل هذه القياسات من مقاييس فترة أو فئوية. ويوجد تخطئة منطقية لمثل هذا النوع من الأدلة. ويعد التشبث بمثل هذه الأدلة نوع من الرهان المشابه لقولنا بأن معظم طلبة التمريض هم من الإناث وإذا كان الطالب الجيد أنثى فقد تكون متخصصة في التمريض، وصيغة أكثر قوة من هذه هو المنحنى الطبيعي نفسه، فمثل هذا الدليل يأخذ بعين الاعتبار حقيقة أن المسافة الفاصلة بين الدرجات الزائية متساوية، ويمكن الحصول على الدرجات الزائية من التحويل الخطي للدرجات الزائية الطبيعية (وهو النوع الوحيد من التحويلات المسموح به لبيانات تدريج الفئة)، ودرجات الاختبار الأصلية هي درجات فئوية.

ويبدو أن ماغنسون (Magenson, 1967, p.13) وافق على هذا الدليل إذ حدد بأن المسافات المتساوية على المتصل والتي يمكن أن تقع على منحنى طبيعي تكون درجاتها متساوية. وعلى هذا يمكننا التعبير عن مواقع الأفراد على المتصل النفسي على تدريج فاصلي. وأحد أوجه قصور هذه المواقع يبدو أنه يجب كتابة عدد كاف من الفقرات واختبار استجابات هذه الفقرات. ويمكننا اختيار مجموعة فرعية من الفقرات حيث يكون توزيع درجاتها قريباً من

التوزيع الطبيعي. وعندما تتطابق فقرات أخرى مفاهيمياً مع السمة فيجب تضمينها في الاختبار للحصول على هذا التوزيع، ومن المشكوك فيه أن هذا صحيح كعلاقة منطقية أو رياضية بين المواقع على المتصل والدرجات الملاحظة عليه. ومشكلة ثانية مع هذا الدليل يكمن في أن المسافات بين الدرجات الزائية قد تكون متساوية عددياً إلا أنها غير متكافئة من وجهة نظر نفسية. فمثلاً ليكن لدينا صف أطفال تتوزع درجاتهم اعتدالياً في اختبار الاستيعاب القرائي. فمن وجهة نظر نفسية فإن المسافة بين قدرات الأطفال الذين درجة أحدهم عند المتوسط وآخر درجته دون المتوسط بدرجة واحدة تكون ذات معنى ومهمة، في حين أن المسافة بين قدرات طفل درجته 3 انحرافات معيارية دون المتوسط وآخر درجته 4 انحرافات معيارية دون المتوسط قد تلاحظ بجهد كبير.

قضية أخرى مرتبطة بمناقشة التدريج الرتبي- الفاصلي في الاختبارات النفسية هو الاختبارات الإحصائية التي يمكن تطبيقها على البيانات. وصندوق باندوا هذا مفتوح للمجتمع النفسي من قبل ستيفنز (1946) عندما دعم بأن: «معظم التدريجات المستخدمة بفعالية وعلى مدى واسع من قبل النفسيين من النوع الرتبي. فالخصائص المتشددة للاحصائيات التي تستخدم المتوسطات والانحرافات المعيارية لا يجوز استخدامها مع هذه التدريجات، فهذه الاحصائيات تتضمن معرفة شيء أكثر من مجرد الترتيب النسبي للبيانات» وبلوّربيورك (Burke, 1963) موقفين لهذه القضية على أنها القياس الموجه والقياس المستقل. وتحتفظ مجموعة القياس المستقل بطبيعة التدريج وهذه تملي نوع من الإجراءات الإحصائية والحسابية التي يمكن تطبيقها على البيانات. فعندما يؤكد علماء الاجتماع عدم اثبات التدريج الفئوي لقياساتهم، فمن الحكمة استخدام الإجراءات الإحصائية واختبارات الدلالة لبيانات التقدير الرتبي. ويتوافر في الاحصاء الاستدلالي مجموعة اختبارات احصائية مجمعة تحت اسم احصائيات لا معلمية (Nonparametric statistics). وهذه الطرائق موضحة في العديد من الكتب مثل (Marasciulo & McSweeney, 1977) .

وتفترض مجموعة القياسات المستقلة أن القياسات التي نحصل عليها تعد أعداداً مثل غيرها (بغض النظر عن التدريج الأصلي)، وتخضع للحسابات والتحليلات الإحصائية التي تحقق افتراضات الطريقة الإحصائية المستخدمة. وتشير هذه المجموعة إلى أن معظم الاختبارات الإحصائية المعلمية لا تتطلب القياس على تدريج فئوي، ولكنها تتطلب افتراضات حول توزيع البيانات، وعندما تحقق البيانات افتراضات التوزيع، فإن هذه المجموعة ترى أنه لا داعي للتضحية بالإمكانات الإحصائية والحسابات المناسبة التي يمكن تحقيقها من خلال الاختبارات الاحصائية المعلمية.. ومن المراجع التي تفضل استخدام القياسات المستقلة يمكن

إيجادها في (Gardner, 1975) وفي (Ware & Benson, 1975). وحديثاً يبدو أن معظم علماء الإجتماع يؤيدون هذه الأفضلية.

وتحديداً، فإن قضية مستوى القياس رتبي أم فاصلي تعد قضية براغماتية. ويجب على مستخدم الاختبار بالأساس أن يسأل: هل يمكن استخدام درجات الاختبار على أنها بيانات تدريج فاصلي؟ افترض أن لديك درجات ستة مفحوصين في اختبار تعيين الفصل التعليمي، وعلى النحو الآتي:

التلميذ	الدرجة
1	37
2	35
3	39
4	22
5	18
6	42

فإذا أراد مستخدم الاختبار تقسيم المفحوصين إلى مجموعتين بالاعتماد على درجات الاختبار، فيُمكنه بسهولة رؤية أن الطلبة (6,3,1) سيكونون في مجموعة واحدة، والطلبة (5,4) في مجموعة أخرى. وتكمن المشكلة في تحديد موقع الطالب (2)، فإذا أشار مستخدم الاختبار إلى وضعه في المجموعة الأولى لأن درجته أقرب إلى درجاتهم، فإن مستخدم الاختبار يكون قد تعامل مع درجات الاختبار على أنها بيانات تدريج فئوي، وأن وضع مستخدم لاختبار الطالب في المجموعة الثانية على أساس ترتيب أعلى ثلاث درجات في مجموعة وأدنى ثلاث درجات في مجموعة أخرى فإنه يكون استخدم درجات الاختبار على أنها بيانات تدريج رتبي. أي الاختيارين أكثر معقولية؟ إن الإجابة على هذا السؤال تتطلب حل تجريبي. ونحن نتفق مع لورد ونوفيك بأن الدرجات إذا كانت تزودنا بمعلومات أكثر فائدة في التنبؤ أو تعيين الصف عند معالجتها على أنها بيانات فئوية، فإنه يجب استخدامها كذلك. ومن جهة أخرى إذا كانت هذه المعالجة تقلل من فائدتها أو مستخدميّتها فإنه يجب استخدام التدريج الرتبي فقط. ويبدو أن هذا الموقف الوحيد الذي يوافق عليه الجميع. وحتى ستيفنز الذي جعل التدريج الرتبي-الفئوي في متناول الجميع تناسى هذا بعد النقد لاستخدام الاحصاء في تحليل بيانات

الاختبارات النفسية. ومن جهة أخرى فإن الاستخدام غير القانوني للإحصاء يمكن أن يوافق عليه عملياً، وفي أمثلة عديدة فإنه أدى إلى نتائج مثمرة.

الخلاصة:

إن تطوير قوانين منتظمة ووحدات معنوية للقياس لتكميم الملاحظات الأولية تعرف بأنها تكوين تدريج ويؤسس قانون التدريج علاقة أو قانون للتناظر بين العناصر في نظام البيانات وعناصر نظام الأعداد الحقيقية. ولنظام الأعداد الحقيقية خصائص الترتيب والمسافات المتساوية بين الوحدات والمركز الثابت. وحالما يتم تأشير الأعداد لعناصر نظام البيانات فإنها تتصف مباشرة بخصائص الأعداد. وقد حدد ستيفنز (1946) أربعة مستويات لتدريج القياس (الاسمي والرتبي والفئوي والنسبي)، وهذه يمكن تمييزها من خلال خصائص الأعداد الحقيقية التي تطبق عليها. فالأعداد في التدريج الاسمي ليس لها معنى الترتيب أو الوحدات المتساوية أو المركز الثابت. والأعداد في النظام الرتبي لها خاصية الترتيب فقط. والتدريج الفئوي له خصائص الترتيب والمسافات المتساوية بين الوحدات ولكن ليس لها مركز ثابت. وتدريج المستوى النسبي له خصائص الأعداد الحقيقية جميعها. ونوقش في هذا الفصل التحويلات المسموح بها لمستويات التدريج المختلفة.

وفي طرائق تطوير التقييم التربوي يجب على الباحث اختيار طريقة تدريج تسمح بالتقييم إلى أقصى حد تسمح به الأعداد لتحديد المواقع على المتصل النفسي الذي له خصائص الترتيب والوحدات المتساوية. وحدد تورغرستون (1958) ثلاث طرائق أساسية لتطوير التدريج هي:

طرائق تتمركز حول الفرد، وطرائق تتمركز حول المثير، وطرائق تتمركز حول الاستجابة. ويركز التدريج المتمركز حول الفرد على موقع الأفراد على المتصل النفسي. وهو أسلوب شائع الاستخدام في الكثير من جهود مطوري الاختبارات النفسية. وينتج عن أساليب التداريج المتمركزة حول المثير والاستجابة فقط نماذج تدريج تسمح للباحث اختبار حسن مطابقة الملاحظات التجريبية والقيم المتنبأ بها الناتجة عن تدريج له خصائص الترتيب والوحدات المتساوية.

وقد تم توضيح أسلوب التدريج التمركز حول المثير بوساطة قانون ثيرستون في الأحكام المقارنة. وخطوات هذه الطريقة هي: 1) إنتاج مصفوفة Pij بالاعتماد على أحكام المفحوصين على أزواج الفترات المحتملة جميعها. 2) تحويل قيم Pij إلى قيم Zij باستخدام جدول المنحنى الطبيعي. 3) اشتقاق مجموعة قيم متنبأ بها لـ Zij. 4) اختبار المطابقة بين Zij و Zij. 5) تحديد مركز مناسب واشتقاق مواقع للفقرات الأخرى على المتصل باستخدام قيم Zij.

وتسمح أساليب التمركز حول الاستجابة بالتدريج الفوري لكل من المثيرات والأفراد، وقد تم استخدام أسلوبي تحليل التدريج لجوتمان وتقنية الإنتشار لكومبس في توضيح هذا الاسلوب. ويستخدم معامل الإنتاج في تحليل التدريج كطريقة لتحديد حسن مطابقة البيانات للنموذج. وفي تدريج كومبس، يؤشر الأفراد إلى ترتيب تفضيلاتهم لعدد محدود من المثيرات، ولأي عدد محدود من المثيرات يوجد أحد أقصى للنماذج المسموح بها لهذه التفضيلات، وقد تم توضيح طريقة لتحديد ترتيب المثيرات على المتصل وتحديد المسافة النسبية بين أزواج المثيرات. وهناك طرائق أكثر تعقيداً من الناحية الاحصائية لاختبار حسن مطابقة نماذج التدريج المتمركز حول الاستجابة، ولكن لم يتم عرضها في هذا الفصل.

وقضية ما إذا كانت القياسات المتمركزة حول الفرد تؤدي إلى بيانات فئوية أو رتيبة تم الكشف عنها من خلال انطباعات ثلاثة: فالأدلة المتعلقة بكون التوزيع الطبيعي للدرجات يؤدي إلى تدريج فئوي لم تكن مقنعة بدرجة كافية. إذ يبدو أن استخدام الاحصاء المعلمي يمكن تعديله بحسب ما تتطلبه الطرائق المستخدمة، حتى عندما لا يكون الباحث متأكداً من أن بياناته لها قوة التدريج الفئوي. وأخيراً، يبدو أن الموقف الحاسم في معالجة البيانات على أنها بيانات تدريج فئوي وذلك عندما يكون بالإمكان تقديم إثبات تجريبي يدعم فائدة الدرجات في التنبؤ أو الوصف من خلال هذه المعالجة.

التمارين:

1/ حدد نوع التحويل المتبع وأنواع البيانات التي تطبق عليها هذه التحويلات:

أ-	ب-	جـ-
75-37	5-12	87-22
53-26	4-15	102-25
109-54	3-18	19-19
57-28	0-21	5-6
65-32	2-24	
	4-15	

2/ استخدم الاستجابات المبينة لعشرة أفراد على 4 فقرات من قبل المستخدمين العامين (أنظر إلى أنماط الاستجابة المسموح بها المبينة في جدول 3-1):

الأفراد	الفقـــرات			
	أ	ب	جـ	د
1	+	+	+	+
2	+	−	+	+
3	+	+	−	−
4	+	−	−	−
5	+	+	−	−
6	−	−	−	−
7	+	−	+	−
8	−	+	+	−
9	+	−	−	−
10	−	+	−	−

أ- رتب الفقرات على أساس عدد الاستجابات الموجبة للفقرة؟

ب- حدد الأفراد الذين أنماط استجاباتهم غير مسموح بها.

جـ- رتب الأفراد الذين استجاباتهم مسموح بها بحيث تكون اتجاهاتهم من الأكثر إلى الأقل نحو الضرب بوساطة المستخدمين العامين.

د- أي الاستجابات على الفقرات تحتاج إلى تغيير بحيث تصبح أنماط الإستجابة أنماطاً مسموح بها؟ كيف يمكن أن تصبح هذه المعلومات مفيدة لمطور الاختبار في مراجعة الفقرات مستقبلاً.

هـ- ما معامل الإنتاجية لمجموعة الفقرات هذه ؟ وكيف يمكن تفسيره؟

د- افترض أن مطور الاختبار كتب 40 فقرة لتحليل التدريج، ما الصعوبات التي يمكن أن يواجهها؟

3/ للموضوعات الآتية: عقاب أطفال المدرسة، استخدام الأسلحة النووية في الحروب، عقاب المجرمين الخارجين عن القانون. حاول تطوير فقرات عددها من 3 إلى 5 مدرجة لأحد هذه المواضيع، وبين إلى ما يجب أخذه بعين الاعتبار لأنماط الاستجابة المسموح بها.

4/ بعد دراسة البيانات الموضحة في جدول (3-2)، أشر إلى المنطقة على المتصل (الاعداد من 1 إلى 7) التي يجب أن يؤشر عليها الفرد ليحصل كل منهم على أنماط الاستجابة المشار إليها.

5/ لكل موقف من مواقف القياس الآتية، وضح أسلوب تطوير التدريج الذي يبدو أكثر ملائمة، أختر أحد الأساليب (التمركز حول الفرد، التمركز حول المثير، التمركز حول الاستجابة):

أ- باحث في التسويق يهدف إلى تقييم إقبال عشر مجلات مختلفة على منتج تجاري واحد.

ب- بروفيسور في كلية يرغب في تطوير اختبار لإعطاء درجات نهائية في مساق الكيمياء.

جـ- يرغب عالم نفس تربوي في تطوير مجموعة مسائل في المهارات الرياضية الأساسية باستخدام مسائل تتزايد في صعوبتها، ويمكنه استخدامها في التسكين التشخيصي للطلاب. ويهدف عالم النفس في اختباره هذا تحديد مستوى الطلبة التدريسي بافتراضه أن الطالب عندما يخطيء في ثلاث مسائل متتالية فإنه لن يجيب عن أي سؤال آخر إجابة صحيحة.

الفصل الرابع

4

خطوات بناء الاختبار

الفصل الرابع

خطوات بناء الاختبار

لاحظ في الفصل السابق أن هدف القياس في العلوم الاجتماعية والتربوية هو وضع الأفراد على متصل كمي وذلك للبناء النفسي المقيس، ويطلق على هذا اسم القياس المتمركز حول الفرد. ويهدف هذا الفصل إلى وصف الخطوة التي تلي في بناء الاختبار أو المقياس المتمركز حول الفرد. ويركز هذا الفصل على الخطوات المنتظمة في بناء الاختبار والمطبقة على مدى واسع من الاختبارات المختلفة. وهذه الخطوات على النحو الآتي:

1/ تحديد الأهداف الأولية التي ستستخدم فيها درجات الاختبار.

2/ تحديد السلوكات التي تمثل البناء النفسي أو نطاقه.

3/ إعداد مجموعة مواصفات للاختبار تصف بدقة نسبة الفقرات الممثلة لكل نوع من السلوكات المحددة في الخطوة الثانية.

4/ بناء ملف أولي من الفقرات.

5/ مراجعة الفقرات وتعديلها عند اللزوم.

6/ تجريب أولي للفقرات ومراجعتها إذا لزم الأمر.

7/ تطبيق الفقرات على عينة كبيرة ممثلة لمجتمع المفحوصين.

8/ تحديد الخصائص الاحصائية لدرجات الاختبار، وعند الضرورة حذف الفقرات التي لا تتفق مع المحكات المحددة مسبقاً.

9/ تصميم وإجراء دراسات الصدق والثبات للصيغة النهائية للاختبار.

10/ إعداد دليل الاختبار. الذي يفيد في التطبيقات، والتصحيح، وتفسير الدرجات (مثل: الجداول المعيارية، واقتراح توصيات مثل معايير الأداء ودرجات القطع وغيرها).

والخطوات السابقة تمثل أقل جهد مطلوب للتأكد من أن درجات الاختبار ذات خصائص تطبيقية تخدم كأداة قياس مفيدة، ويضمّن العديد من المتخصصين في بناء الاختبارات خطوات إضافية إلى التسلسل المذكور أعلاه أو أنهم يكرروا بعض الخطوات مرات عديدة.

ويهدف التسلسل المطروح هنا تقديم دليل للمبتدئين في بناء الاختبارات وتطويرها عند استخدامها في عمليات التقويم أو إجراء الدراسات. فالخطوات الست الأولى تعد جوهرية في إعداد الصورة الأولية للاختبار وهي الهدف الرئيس لهذا الفصل. وأما الخطوات المتبقية (من 7 إلى 10) فهي الموضوعات الرئيسة للفصول التالية في هذا الكتاب.

الخطوة الأولى: تحديد الأهداف التي ستستخدم فيها درجات الاختبار:

يجب أن تأخذ الخطوات المنتظمة في بناء الاختبار بعين الاعتبار الأهداف الأساسية التي ستستخدم فيها درجات الاختبار. فعلى سبيل المثال افترض أنه تم تكليف خبير في القراءة تطوير كتاب في الاستيعاب القرائي للطلبة الذين سينخرطون في التدريس الجامعي ومن المؤكد أن درجات مثل هذا الاختبار ستستخدم في إتخاذ قرارات تتعلق بالقبول والتسكين والتشخيص. ومن غير المؤكد أن تطوير اختبار واحد يمكن أن يحقق هذه الأهداف جميعاً بدرجة مثالية. وكما سنرى في الفصل الخامس فإن الاختبار الذي سيميز بين المفحوصين على مدى واسع من القدرة (أو الاختبار ذو الحساسية العالية) يجب أن يتألف من فقرات متوسطة الصعوبة، وذلك لتعظيم (تضخيم) تباين درجات المفحوصين إلى أكبر قدر ممكن. ومن جهة أخرى، يستخدم الاختبار التشخيصي في تحديد المجالات التي يكمن فيها الضعف عند الطلبة ذوي القدرة المتدنية، ومثل هذا الاختبار يجب أن يتضمن عدد كبير وأساسي من الفقرات السهلة نسبياً لمجتمع المفحوصين. كذلك فإن الاختبار المصمم لتقييم الحد الأدنى من الكفاءة يختلف عن الاختبار المصمم لاختيار المتنافسين في برنامج تعليمي أو تربوي. لذلك فإن توضيح الأهداف الرئيسة التي ستستخدم درجات الاختبار من أجل تحقيقها، وتحديد الأولويات للاستخدامات المتعددة للاختبار ترجح أهمية وفائدة الاختبار للهدف الأساسي الذي أُعد الاختبار من أجله.

الخطوة الثانية: تحديد السلوكات التي تمثل البناء

كقاعدة عامة تبقى الخطوات العملية التي يتم خلالها تحويل البناء النفسي إلى مجموعة فقرات خاصة وغير رسمية وغير موثقة وقد ناقش كل من كروبناخ (1970)، ورويد وهالادينا (1980)، وشوميكر (1975) هذه الخطوات مع تأكيدهم على ذلك خاصة عندما يتعلق الأمر بالاختبارات التحصيلية. هذا ويُجري مطوِّر الاختبار في العادة تحليلاً مفاهيمياً لواحد أو

أكثر من أنماط السلوك التي يعتقد أنها تمثل البناء المستهدف، ومن ثم يحاول إعداد فقرات تغطي هذه السلوكات، ولكن مثل هذا الإجراء قد ينتج عنه حذف مجالات مهمة من السلوك أو احتوائه لمجالات تكون واضحة فقط في عقل مطوِّر الاختبار، ومثل هذا الإسلوب ينتج عنه ذاتية عالية وتعريف خاص مزاجي للبناء. وللحصول على صورة واضحة منقحة وشاملة للبناء المراد قياسه على مطور الاختبار توظيف واحدة أو أكثر من الأنشطة الآتية:

1/ **تحليل المحتوى**: وفي هذا الأسلوب يواجه الأفراد بأسئلة ذات نهاية مفتوحة حول البناء المراد قياسه، ومن ثم تصنيف استجاباتهم ضمن فئات معينة، والفئات التي تتكرر بكثرة تُعد عناصر رئيسية للبناء. وقد نشر جيرسلد (1952) نتائج تحليل محتوى الإنشاء وذلك من خلال وصف الأطفال لأنفسهم، واعتمد الفئات الناتجة عن هذا الوصف أساساً في بناء فقرات استبانتين واسعتي الانتشار صممتا لقياس مفهوم الذات عند الأطفال.

2/ **مراجعة الأبحاث**: فالسلوكات التي تم دراستها بشكل متكرر من قبل الآخرين تستخدم في تحديد البناء المراد قياسه. وعلى مطوِّر الاختبار أن يختار إحدى الطريقتين السابقتين أو أن يختار ما قدمه أحد المنظرين المعروفين في تحديد الفئات السلوكية التي يجب تمثيلها في فقرات الاختبار.

3/ **الأحداث العرضية الحرجة**: وهنا تحدد قائمة السلوكات التي تميز الأداءات المتطرفة على متصل البناء. ويعزى هذا الاسلوب إلى فلانجان (Flangan, 1954)، إذ قام بسؤال المشرفين على الوظائف عن وصف المواقف التي يقوم المستخدمين بتوظيفها، وكان لها تأثيرٌ واضحٌ سواء أكان إيجابياً أم سلبياً، ومنها أعدّ قائمة للسلوكات الحرجة استخدمت في إعداد تدريج (متصل) للأداء الوظيفي.

4/ **الملاحظات المباشرة**: هنا يحدد مطوِّر الاختبار السلوكات من خلال الملاحظة المباشرة، فمثلاً يطوِّر المرشد المهني استبانة لتقييم الضغط الوظيفي في المواقف الخطرة جداً، وقد يجد مطوِّر الاختبار الملاحظات الواقعية لمثل هؤلاء مساعداً في تحديد المواقف التي تعد مصادر الجهد في الضغط الوظيفي.

5/ **أحكام الخبراء**: يمكن لمطوِّر الاختبار الحصول على مدخلات ممن هم خبراء أوائل في البناء نفسه، وذلك من خلال استبانات مكتوبة أو لقاءات شخصية لجمع المعلومات. ومثال ذلك يمكن للمتخصص في علم نفس الشخصية أن يطوِّر قائمة لتدريج أداء فريق التمريض في مستشفى كبير من خلال تغطية أو مسح لمجموعة المشرفين على التمريض يتم من خلالها تحديد أنواع الأداءات التي يجب أن تتضمنها.

6/ الأهداف التدريسية: لتطوير اختبار تحصيلي يتم سؤال الخبراء في الموضوع لمراجعة المحتوى وكذلك لتطوير مجموعة من الأهداف التدريسية المتعلقة بالموضوع نفسه. فالأهداف التعليمية تخصص السلوك الملاحظ الذي يمكن أن يؤديه الطالب بعد إتمامه للمساق التدريسي، ومثل هذه الأهداف تعمل كشبكة وصل لكل من المحتوى الذي يجب التركيز عليه وطبيعة المهام التي يجب أن يكون المفحوص قادراً على أدائها. والوصف التفصيلي لعملية تطوير الأهداف متوافرة في معظم كتب القياس مثل بوبام (Popham, 1981) وبراون (Brown, 1983) وميرنز وليهمان (Mehrens & Lehman, 1984).

معاينة المجال Domain Sampling

هنالك بناءات نفسية يكون الاهتمام بها أولي مثل الذكاء والإبداع والتطور الخلقي، إذ أن الأفراد يختلفون في كميات هذه الخصائص التي يمتلكونها. ولتطوير اختبارات عادةً تجري عملية تحليل مفاهيمي لعناصر السلوك الرئيسة للبناء، ثم إعداد فقرات لعناصر السلوك هذه، وأخيراً اختيار الفقرات التي تُبرز التباين المتوقع في الأداء. فالفقرات ذات التباين المنخفض تحذف من الصيغة النهائية لأنها لا تمثل البناء بشكل مناسب. ولأن درجات الاختبارات المبنية بهذه الطريقة تفسر بمقارنة أداء الفرد بأداءات الآخرين، لذا يطلق على هذه الاختبارات اسم معيارية المرجع.

كذلك، ففي مواقف عدة يلزم قياس أداء المفحوصين بمستويات أكثر تجريداً كما هو الحال في الاختبارات التحصيلية، وذلك عندما يهدف مستخدم الاختبار التأكد من امتلاك المفحوص حد أدنى من الكفاية في موضوع معين أو عند تقويم فاعلية برنامج تعليمي. واقترح إيبل (Ebel, 1962) في مثل هذه الحالات الحاجة إلى درجة معيارية للمحتوى (Content standard score) تكون ذات معنى مفسر للمفحوص بغض النظر عن أداء الآخرين على الاختبار، وقد اقترح جليسر (Gleser, 1963) استخدام مصطلح القياس محكي المرجع على درجات الاختبار التي تشتق معناها من المستويات المطلقة للمفحوصين من خلال استجاباتهم على سلسلة فقرات تكافيء كل مجموعة منها مستوى معروف من الكفاية وذلك على محك الأهمية (Criter-ion of importance). والاسلوب المذكور سابقاً في أجرأة البناء لتطوير اختبارات للأهداف المحكية غير كافٍ. وعلى الأغلب يستخدم مطور الاختبارات التحصيلية محكية المرجع مجموعة من الأهداف التدريسية يحدد من خلالها مجالات الأداء التي يمكنه من خلالها تكوين استدلالات من خلال درجات الاختبار، ويطلق على مجال الأداء هذا إسم مجال الفقرات (Item domain). ويعد مجال الفقرة مجتمعاً محدداً جيداً من الفقرات يمكن من خلالها بناء صيغة اختبارية أو أكثر وذلك باختيار عينة من الفقرات من هذا المجتمع، أو يطلق على هذا أحياناً أسلوب معاينة المجال في بناء الاختبار.

ومن الواضح، صعوبة توليد الفقرات الممكنة جميعها للمجال، لذا فإن استخدام عدد قليل منها يُعد اقتصادياً لكنه غير مُجدٍ في مجالات عديدة. وكبديل لإنتاج مواصفات فقرات المجال والتي تعد بصورة تجريدية ومنها كتابة الفقرات على وفق هذه المواصفات هنالك مراجعة واسعة مكثفة للأساليب المقترحة لتحديد مجال الفقرات ثم كتابتها عرضها كل من ميلمان (Millman, 1974)، ورويد وهالادينا (Royd & Haladyna, 1980)، وشوميكر (Shoemaker, 1975). ومن الأساليب واسعة الإنتشار ما يُعرف باسم مواصفات الفقرات (Iternspecification) أو ما يسمى بالهدف الموسع أو المضخم (amplified objective). وخصائص الفقرات كما حدد بوبام (Popham, 1974) تتضمن مصادر محتوى الفقرات، ووصف الموقف المثير أو المُشكِل، وخصائص الإجابة الصحيحة، وفي حالة فقرات الاختيار من متعدد خصائص الاستجابات الخاطئة، والشكل (4-1) يبين مثال توضيحي.

المهارة الفرعية Subskill: يظهر قدرة (يبين قدرته) في ضرب الكسور العشرية.

توصيف الفقرات Item descriptor: ضرب الكسور العشرية: الحساب.

خصائص المثير:	**خصائص الإستجابة**
1/ تتطلب المسألة ضرب كسرين عشريين، كسور و/أو كلاهما.	1/ الصيغ كل الاختيارات العددية يجب أن ترتب حسب المنزلة العشرية وترتب إما تصاعدياً أو تنازلياً.
2/ المسألة يجب أن تكتب في جملة أو صيغة أفقية، وعند استخدام الصيغة الأفقية، التعليمات يجب أن تطلب: جد الناتج أو اضرب.	2/ عرض أربعة بدائل للإجابة: أ- الإجابة الصحيحة. ب- إحدى الإجابات الخاطئة يجب أن تعكس خطأ في إعادة التجميع أثناء عملية الضرب.
3/ أحد العوامل (الإعداد) يجب أن يتكون من ثلاثة منازل غير صفرية، والآخر يجب أن يتكون من ثلاثة منازل أثنان منها غير صفرية أكبر من 5.	جـ- إحدى الإجابات الخاطئة يجب أن تعكس خطأ ترتيب أثناء عملية الضرب.
4/ كلا العددين يجب أن يتضمنا على الأقل منزلة عشرية واحدة.	د- إحدى الإجابات الخاطئة يجب أن تعكس إما (a) حذف إشارة الكسر العشري أو (b) خطأ في تحديد موقعها.
5/ يجب أن لا تتضمن الإجابة أكثر من 4 منازل عشرية.	3/ إجابة أخرى ممكنة «لا شيء مما ذكر» يمكن أن يحل محل أ أو ب أو جـ. وهذا البديل يجب أن يظهر في البديل الرابع (د).
6/ من الضروري إجراء عملية إعادة تجميع (re-grouping) مرتين على الأقل.	
7/ عند اختيار العوامل (factors) لا يجب استخدام الكسر العشري أكثر من مرتين.	

شكل (4-1). خصائص فقرات اختبار في المهارات الرياضية الأساسية.

مقتبس من منشورات فلوريدا في اختبار المعلمين 2-1 استخدم بإذن من قسم التربية في جامعة فلوريدا.

ويكون لاستخدام خصائص الفقرات مزايا جيدة خاصة عند إعداد ملف اختباري كبير، وتراجع فقراته من قبل عدد من المختصين في كتابة الفقرات، وعندما يكون هؤلاء ضالعين في توصيف الفقرات يمكن توليد عدد كبير من الفقرات المتوازية التي تقيس هدف معين وبفترة وجيزة. وعلاوة على ذلك فإن استخدام عدد من المختصين في كتابة الفقرات لفترة زمنية غير قصيرة فإن مواصفات الفقرات تؤكد على أن صيغ الفقرات الحالية تكافيء تلك المستخدمة في صيغ سابقة.

وهناك طريقة أخرى بتركيب أفضل لخصائص الفقرات إما من خلال الفقرات العددية (item algorithm) أو صيغ الفقرات (item forms). ويطبق مصطلح الفقرات العددية على الأغلب عندما يكون محتوى الفقرات كمياً. بينما يستخدم المصطلح صيغ الفقرات عندما يكون محتوى الفقرات لفظياً. لتوضيح الفقرة العددية، تفحص الهدف:

– أن يتمكن الطالب من حساب الفرق بين أي عددين موجبين في الصيغة (أ – ب = ؟) عندما يكون أ > ب. هذا ويمكن تكوين عدد كبير من الفقرات بهذه الصيغة، وذلك بتعويض قيم مختلفة لكل من (أ و ب)، ومجال الإجابة قد يكون عدداً واحداً أو مجموعة متنوعة من الأعداد.

وللمحتوى اللفظي، تكتب الصيغة العامة للفقرات حيث يكون لها تركيب ثابت، لكن أرومة السؤال تتضمن واحداً أو أكثر من العناصر المتغيرة. وتحدد هذه الصيغة فئة كلية من الفقرات اعتماداً على مجموعة من عناصر الإحلال، وتعويضها في العنصر المعرفي لأرومة الفقرة. وأمثلة من هذه الفقرات:

عندما يظهر على شجرة الحمضيات أ، فمن المحتمل أن هذا من أعراض ـــــــ:

1. نقص الغذاء. 2. ضرر عشبي.

3. ضرر من الطقس البارد. 4. عدوى فيروسية.

5. عدوى بكتيرية.

وفي المجموعة (أ) يمكن كتابة أعراض تظهر على الورق واللحاء مثل اصفرار الورق، تشقق اللحاء، بقع صدأ على الأوراق، سناج أسود على الأوراق، تقشر اللحاء، داء الفسيفساء أو ترقيط الورق بألوان مختلفة، تقرح وتآكل اللحاء، والإجابة الصحيحة لهذه الفقرة تتغير تبعاً للعنصر (أ) الذي سيستخدم في أرومة السؤال.

وأسلوب أكثر بنيوية more structured لخصائص النطاق ومشابه لاستخدام صيغة الفقرة يستخدم جمل خرائطية (mapping sentences) التي تحدد مسبقاً أجزاء الكلام أو اللغة (speech)، وأشباه الجمل، والشروط التي تتغير بشكل منتظم والتي تتحدد بمواقع معينة من الجملة ضمن تركيبها أو بناءها. وقد اقترح جوتمان (Guttman, 1969) استخدام الجمل الخرائطية كجزء من نظريته (facet theory) وذلك في تحديده لخصائص الفقرات في الاختبارات المعرفية أو إستبانات الإتجاهات التي قد تؤثر على نمط الإستجابة. وناقش كل من بيرك (Berk, 1978) وميلمان (Millman, 1974) تطبيق الجمل الخرائطية في تعريفهم للنطاق. واقترح بورمث (Bormuth, 1970) أسلوباً في تحويل القواعد عند بناء فقرات الاختبار بالإستناد إلى مقالة مكتوبة، وهذه تشبه أسلوب الفقرة- الصيغة (item-forms method).

وعند الأخذ بعين الإعتبار مستوى التخصيص عند تحديد النطاق في موقف معين يجب أن تؤخذ النقاط الآتية بعين الاعتبار:

أولاً: معنى الدرجات في الاختبار محكي المرجع يعتمد على (1) تحديد نطاق السلوك في الأهمية التطبيقية (2) بناء فقرات بطريقة تسمح لمستخدم الاختبار الاستدلال بأن الأداء على هذه الفقرات يمثل الأداء على نطاق السلوك الكلي للفقرات. ومن الواضح هنا أنه كلما كان النطاق محدداً بشكل واضح وجلي وتخلي معد الفقرات عن الذاتية في اختياره لصيغ الفقرات ومحتواها كلما كانت ثقة مستخدم الاختبار بأن الفقرات المستخدمة في صيغة معينة للاختبار تمثل النطاق (بمعنى أنه عند بناء صيغة أخرى للاختبار فإنها تكون ممثلة للاختبار الأول بشكل دقيق). وعندما يصبح النطاق محصوراً ومحدداً فإن إمكانية التعميم وبثقة على النطاق المتضمن لمجال محدد جداً من السلوك يكون مناسباً جداً في بعض القياسات، ولكن يكون عليه تساؤلات في مجالات قياسية أخرى.

ثانياً: قضية مرتبطة معتمدة على تخصيص الفقرات في خفض الذاتية عند استخدامها في اختيار صيغ الفقرات ومحتواها تحل ببساطة محل إتخاذ قرار ذاتي في كتابة الفقرات إلى الحد الذي تم تخصيص الفقرات فيه.

أخيراً، إذا كانت نتائج تخصيص معين للفقرات أدت إلى إنتاج فقرات مطابقة لفقرات ضعيفة تقنياً أو متأثرة بمجموعة الاستجابات، فإن الاختبار يكون متضمناً نسبة كبيرة من الفقرات الخاطئة والتي قد تظهر فيما لو أن كتبة الفقرات لم يلتزموا في اتباعهم لمواصفات الفقرات. ومن المهم أيضاً مراجعة خصائص الفقرات قبل استعمالها، وعند تطوير فقرات عديدة لكل توصيف للفقرات على أسس استدلالية، إذ يصبح اكتشاف الضعف أو التصدع في الفقرات ممكناً.

إعداد مواصفات الاختبار Preparing Test Spercification

بعدما يتم اختيار الأهداف وخصائص الفقرات وفئات السلوك الأخرى فإن مطور الاختبار يحتاج إلى صياغة خطة لاتخاذ قرار بالتأكيد على أن كل العناصر متضمنة في الاختبار. وبالتحديد يجب أن يكون هنالك توازناً للفقرات الممثلة لعناصر الاختبار بنسبة أهمية ذلك العنصر من وجهة نظر مطور الاختبار. وعندما يكون نطاق الفقرات محدداً من خلال مجموعة خصائص الفقرات تصبح القضية: ما عدد الفقرات التي يجب أن تمثل كل تخصيص للفقرات. وبدون وجود خطة معينة فإن الصيغ المتتابعة للاختبار والناتجة عن الخصائص نفسها للفقرات قد تكون مختلفة جوهرياً عندما يؤكد معدو الفقرات على بعض الخصائص دون غيرها، كذلك فإن المجالات التي يتم التأكيد عليها قد تختلف من صيغة لأخرى. وفي حالة أكثر شيوعاً، عندما لا يكون هنالك إعداداً لخصائص الفقرات فإن تطوير خصائص الاختبار يتطلب من مطور الاختبار أن يهتم بخاصيتين للفقرات وباستقلالية لكل منهما عن الأخرى- المحتوى والعمليات العقلية أو الإجرائية التي يستخدمها المفحوص في حل المهمة المطلوبة في الفقرة. وهذا مهم جداً خاصة في تطوير الاختبارات التحصيلية، وكما هو مبين في المثال الآتي:

افترض الأهداف الآتية في الهندسة المستوية وفي الوحدة التعليمية نفسها.

أ/ حدد المصطلحات الأساسية المرتبطة بالدائرة (مثل نصف القطر، المحيط، الزاوية المركزية).

ب/ إحسب قياس المساحات، المسافات، محيط الدوائر، الزوايا التي تقاس باستخدام خصائص الدوائر.

من الواضح أن الهدف الأول يتطلب استرجاع المعلومات المخزونة في الذاكرة، في حين أن الهدف الثاني يتطلب معرفة هذه المفاهيم وتطبيق المباديء التي تحدد العلاقة بين أثنين أو أكثر من هذه المفاهيم. لذلك فإن التخصيص البسيط للمحتوى ودون الإشارة إلى مستويات أو أنواع العمليات المعرفية التي يجب تمثيلها في فقرات الاختبار، لا يزودنا بدليل مناسب في بناء الاختبار.

وقد اقترح العديد من المؤلفين نظاماً هرمياً في تصنيف العمليات المعرفية التي تعد مهمة في تطوير خصائص الاختبار. والأكثر شيوعاً وشهرة في هذه التصنيفات تصنيف بلوم (Bloom, 1956) والمؤلف من الفئات الرئيسة الآتية:

1/ المعرفة- استرجاع الحقائق المذكورة في المادة التعليمية بصيغة مشابهة للطريقة التي عرضت أثناء التعليم، ومثال لهدف بهذا المستوى: تسمية عواصم بعض الولايات.

2/ **الفهم أو الإستيعاب** – الترجمة، التفسير أو استقراء المفاهيم بطريقة مختلفة عن الطريقة التي عرضت بها، وهدف من هذا المستوى يتطلب تمييز الأسماء في جمل لم تستخدم في الأمثلة المطروحة في الصف، أو تقديم مثال لعدد أولي غير الذي عرض في الصف أو الكتاب المدرسي.

3/ **التطبيق**– حل مشكلات جديدة من خلال استعمال مبادئ أو تعميمات مألوفة. مثل حساب مقاومة موصل كهربائي باستخدام قانون أوم، وذلك عندما لا يبين السؤال القانون اللازم لحل المشكلة.

4/ **التحليل**– تحليل المعلومات أو المشكلة إلى عناصرها باستخدام عملية تتطلب معرفة عناصر متعددة و/أو مباديء تنظيمية و/أو عبر هذه العناصر. مثل تحديد جنس عينة جديدة من النبات من خلال خصائص أوراقها وأزهارها.

5/ **التركيب**– جمع العناصر في وحدة باستخدام البناء الأصلي أو حل مشكلة تتطلب اتحاد مجموعة من المباديء بتسلسل في موقف روائي. مثل كتابة برنامج محوسب يجري حسابات لمجموعة بيانات مسجلة باستخدام مدخلات، مخرجات، حلقة، تحويلات منطقية لجمل بتسلسل فعّال في الإجراءات.

6/ **التقويم**– وذلك باستخدام محك (يولده ذاتياً) داخلي أو خارجي لإصدار أحكام بلغة واضحة ومنطقية من وجهة نظر فلسفية أو أدبية. مثل كتابة مراجعة نقدية لمقالة صحفية تصف بحث أولي في علم النفس الاجتماعي أو الشخصي.

وقد زودنا بلوم وزملاؤه (Bloom, etal, 1971) بتفسير إضافي وأمثلة لهذا التصنيف. والأمثلة الإضافية لتصنيف العمليات المطلوبة في اختبارات الأداء الأمثل– أقصى أداء – التي وصفت من قبل إيبل (Ebel, 1972) واقترح فيها عمليات المعرفة واستخدام مصطلحات لاسترجاع حقائق المادة التعليمية، واستخدام التعميمات والتفسير والحساب وتنبؤ النتاجات والتوصيات. كذلك قدم ثورندايك وهاجن (Thorndike & Hagen, 1977) تفصيلاً لفئات تمييز المصطلحات والألفاظ، وتحديد الحقائق، وتحديد المباديء، والمفاهيم والتعميمات، وتقويم المعلومات، وتطبيق المباديء في المواقف الجديدة أو غير المألوفة. وفي الأغلب فإن معد الاختبار يختار عدداً قليلاً من العمليات المذكورة في واحد من التصنيفات أو أنه يجمع بين فئات من أكثر من تصنيف في تغطية الحاجات الخاصة بالاختبارات.

وبسبب التنوع الكبير في مستوى العمليات المعرفية التي تظهر في الفقرات التي تغطي الموضوع الواحد فإن خصائص الاختبار تقدم إرشادات مهمة لمعد الاختبار من خلال مستوى

العمليات التي يجب تمثيلها في الاختبار. كذلك فإنها تؤشر للتوازن الذي يجب الحفاظ عليه في اختيار العمليات المختلفة في الاختبار ككل. وأحد التراكيب المهمة لمواصفات الاختبار التي تأخذ بعين الاعتبار كلاً من المحتوى والعمليات يُعرف بجدول المواصفات. ويتألف هذا الجدول من بُعدين على الأقل، يتضمن الأول مجالات المحتوى الرئيسة والآخر العمليات العقلية. ويبين الشكل (4-2) جدول توضيحي لمواصفات اختبار فرعي لامتحان شهادة كفاءة المعلم. (هذا المثال يوضح مستويات عقلية مختارة من أكثر من تصنيف على الأرجح). والعدد الموجود في كل خلية يُعد الوزن المناسب المؤشر للأهمية النسبية الذي يؤمل أن يكون ضمن المحتوى والعملية المرتبطة بتلك الخلية. ففي الاختبارات التحصيلية تعكس الأوزان في الخلايا الزمن اللازم للسيطرة على المادة التعليمية أثناء تدريس المساق. ويجب أن يكون مجموع أوزان الخلايا مساوياً (100). وكل وزن يؤشر نسبة الفقرات أو الدرجات في الاختبار المخصصة للمحتوى والعملية. ومثل هذه الأوزان تحدد عدد الفقرات التي يجب كتابتها لتمثل الخلية بشكل مناسب بالنسبة للمجال الذي نهتم به. وفي الشكل (4-1) فإن 1% من الفقرات يجب أن تقيس استرجاع مبادىء تنظيم الغرفة الصفية، و 10% يجب أن تقيس تطبيق مبادىء تنظيم الغرفة الصفية. كذلك يزودنا جدول المواصفات بطريقة مناسبة لوصف الاختبار للمهتمين باستعماله وتطبيقه. وعند جمع النسب في كل عمود نجد أن 7% من فقرات الاختبار تقيس مهارات مستوى المعرفة و 62% منها تقيس مهارات التطبيق و 31% منها تقيس مهارات حل المشكلات. وبجمع النسب في كل صف نستخلص النسبة المئوية الكلية للفقرات التي تغطي فئة كل محتوى على التوالي. وعندما يقرر معد الاختبار عدد فقرات الاختبار الكلي فإن عدد فقرات كل خلية التي تحقق جدول مواصفات الاختبار تساوي وزن الخلية مضروباً بعدد الفقرات الكلي.

وفي استبانات الاتجاهات أو الأداء الملاحظ فإن معد الاختبار يستخدم سلوكات من المجال الإنفعالي في إعداد جدول مواصفات الاختبار. وقد قدم كل من كراثول وزملاؤه (Krathwohl, etal., 1964) تصنيفاً قد يكون مفيداً لهذه الأهداف، وتجد في مؤلف بلوم وزملائه (1971) وصفاً تفصيلياً لمستويات التصنيف الخمسة التي اقترحوها وأمثلة على استعمالاتها في تطوير الاختبار.

مجالات المحتوى / العمليات العقلية	المستوى			المجموع
	المعرفة	التطبيق	حل المشكلات	
1. إدارة الصف الكفايات 6 12 13 15 16 17 20 22	1	10	9	20
2. في تطور الطلبة الكفايات 6 10 11 12 16 17 20 21 22 23	1	11	11	23
3. تقويم، تسجيل، كتابة تقارير عن تقدم الطلبة الكفايات 6 7 10 12 14 18	2	16	1	19
4. المواد التعليمية الكفايات 10 12 13 15 21	1	6	2	9
5. الأهداف التدريسية الكفايات 9 10 11 12 14		7	3	10
6. التعلم والتعليم الكفايات 6 9 10 11 12 13 15 17 21 22 23	2	12	5	19
المجموع	7	62	31	100

شكل (2-4): جدول مواصفات اختبار فلوريدا لكفاءة المعلم

مقتبس من منشورات اختبار فلوريدا لكفاءة المعلم 2-1 Teacher Certification Examination .

استخدمت بإذن من قسم التربوي في جامعة فلوريدا Florida department of Educ.

بناء الفقرات Item Construction

وصف ليندكويست (Lind quist, 1936) مهمة مطور الاختبار بأنها تتطلب نوعين رئيسيين من القرارات المهمة – ماذا ستقيس وكيف ستقيسه؟ وخلال بناء الاختبار يجب عنونة النوع الأخير من القرارات، ويتضمن بناء مجموعة من الفقرات التي تقيس بناءاً معنياً الفعاليات الآتية:

1/ اختيار صيغة مناسبة للفقرات.

2/ إثبات أن الصيغة المقترحة ملائمة لفئة المفحوصين.

3/ اختيار وتدريب معدّي الفقرات.

4/ كتابة الفقرات.

5/ مراقبة تقدم معدّي الفقرات ونوعية الفقرات.

ويتم تحديد تركيب الفقرات وصيغتها مباشرة فيما إذا عمل معدي الفقرات من خلال صياغة الفقرات أو مواصفاتها. وعندما لا تستخدم مواصفات الاختبار يبقى من المهم اتخاذ قرارات حول صيغ الفقرات عند بناء الفقرات عوضاً عن تركها لأذواق معدي الفقرات الخاصة. وعند إتحاذ قرار حول صيغة معينة للفقرات يؤمل من مطور الاختبار مراجعة الأدوات المشابهة في المجال، ودراسة التقارير المعدة عن تطويرها وقد يكون آراء الخبراء مساعداً في تقرير مثل هذه الأمور من مثل ما إذا كان المفحوصون مؤهلين لدرجة كافية ليتقدموا لاختبارات جمعية بالورقة والقلم، وما إذا كان المستخدمون في الشركات يستغرقون وقتاً في كتابة إجابات الأسئلة ذات النهاية المفتوحة أو ما إذا كان طلبة الصف الثالث يمكنهم التمييز بين خمسة نقاط على متصل موافق – غير موافق. وفي الوقت نفسه من الضروري جمع بيانات على مدى ضيق باستخدام عدد قليل من نموذج الفقرات الأصلي للإجابة على مثل هذه الأسئلة وذلك قبل إعداد عدد كبير من الفقرات بالصيغة نفسها.

وحالما يتم اختيار صيغة معينة من الفقرات فإن مطورٌ الاختبار يجب عليه مراجعة المصادر القياسية –المعيارية– في كتابة الفقرات ليأخذ بعين الاعتبار المقترحات اللازمة في كتابة فقرات من هذا النوع. وقائمة هذه الإرشادات يجب إعدادها وتوزيعها على كتبة الفقرات خاصة عندما يكونوا من غير المختصين بالموضوع.

صيغ فقرات اختبارات أقصى أداء: Item Formats for Optimal Performance Tests

لاختبارات أقصى أداء (المعروفة باختبارات الاستعدادات أو التحصيل) يمكن اعتماد مدىً

واسع من صيغ الفقرات. وقد قسم بوبام هذه الصيغ إلى فئتين رئيستين - تلك التي تتطلب من المفحوص كتابة الإجابة (صياغتها) (مثل أسئلة المقال أو الإجابات القصيرة) وتلك التي تقدم للمفحوص استجابتين أو أكثر وتتطلب من المفحوص اختيار إحداها. ولأن الأخيرة يمكن تصحيحها بذاتية ضعيفة فإنه يطلق عليها إسم فقرات اختبار موضوعية، والصيغ الثلاث الأكثر إنتشاراً في الأسئلة الموضوعية هي:

1- اختيار البديل: Alternate Choice

جملة أو سؤال لها استجابتان محتملتان (صح - خطأ) أو (نعم - لا).

مثال: السبب الأكثر احتمالية للسخام الأسود الذي يغطي ورق أشجار الليمون هو غزو الحشرات (صح - خطأ).

2- اختيار من متعدد: Multiple Choice

أرومة يقدم فيها السؤال أو تطرح مشكلة، والإجابة الصحيحة أو مفتاح الإجابة مع أثنين أو أكثر من الاستجابات الخاطئة التي تسمى أخطاء -foils-.

مثال: شجرة ليمون فيها أوراق عديدة مغلفة بالسخام الأسود. قد يكون هذا بسبب:

أ- ضرر عشبي	ب- عدوى بكتيرية
جـ- هجوم الحشرات	د- نقص الغذاء

3- المزاوجة: Matching

تستعمل جملة المبدأ في ربط الأشياء من قائمتين منفصلتين، قائمة من المثيرات أو المقدمة المنطقية، وقائمة من الإستجابات.

مثال: لكل علاقة مذكورة في القائمة اليمنى، سجل رمز السبب الأكثر احتمالاً من القائمة اليسرى:

العلامة	السبب المحتمل
1- أوراق صفراء	أ. التعرض للطقس البارد
2- تشقق اللحاء	ب. نقص النيتروجين
3- بقع صدأ على الأوراق	جـ- عدوى فطرية
4- أوراق مرقشة (مبقعة)	د- ضرر عشبي
	هـ- عدوى فيروسية

ومعظم كتب مقدمة في الاختبارات والقياس تقدم من خلال فصول عدة مقترحات لكتابة مثل هذه الأسئلة مثل (Brown, 1983)، و (Mehrens & Lehman, 1984)، و (Popham, 1981)، و (Sax, 1980). ونقطة مهمة في صياغة الأسئلة الموضوعية بأشكالها كافة هو أن تظهر جميع الإستجابات مقبولة منطقياً من قبل المفحوصين الذين لديهم المعرفة أو المهارة التي صمم الاختبار لقياسها. وكمثال في فقرة الاختيار من متعدد الأرومة المكتوبة جيداً يتبعها الإجابة الصحيحة وثلاثة بدائل خاطئة التي لا تقيس معرفة المفحوص في المجال المطلوب، ومعظم المفحوصين لديهم القدرة على حذفها جميعاً عدا الإجابة الصحيحة باستخدام الفطرة، وكنتيجة تنخفض قيمة الفقرة. وللتغلب على مثل هذه المشكلة فإنه يتم بناء البدائل الخاطئة من خلال الأخطاء المفاهيمية، والتفسيرات الخاطئة، أو الناتجة عن أخطاء حسابية. وهنا على مطور الاختبار أن يتذكر بأن بناء الفقرة الجيدة يتطلب ليس معرفة المحتوى فقط ولكن معرفته بمجتمع المفحوصين للتأكد من أن الاختيارات الخاطئة تبدو معقولة.

وكقاعدة عامة، ينصح مطوّروا الاختبارات إعطاء أهمية لاختيار صيغة الفقرات الملائمة لحاجات المفحوصين، وتجنب الصيغ غير المألوفة أو غير المجربة، ودون أن يكون هنالك مناداة باستخدامها. وحديثاً، ينادي البعض بالتحول عن صيغة أسئلة الاختيار من متعدد ذات الشعبية العالية، وفي الوقت نفسه هنالك اتّجاهٌ ينادي باستخدامها من جهة أخرى. فمن جهة نادى إيبل (Ebel, 1982) بالاستخدام الواسع لفقرات ببدائل أكثر تركيباً. كمثال : كثافة الثلج 1– أكبر 2– أقل من كثافة الماء. ووجهة نظره أنه في مجالات عديدة يمكن التعبير عن المعرفة على شكل سلسلة من العلاقات الوظيفية أو المبادىء، وأن كل منها يمكن أن يكون الأساس لفقرة أكثر بناءاً أو تركيباً. مع ذلك يبدو أن كل فقرة تستطيع اختبار قدر قليل من المعلومات. وقد ناقش إيبل كل من:

1/ يمكن طرح عدد كبير من هذه الأسئلة في وقت قصير، مما يسمح بمعاينة قدر أكبر من محتوى النطاق في الاختبار.

2/ الأداء على مثل هذه الفقرات يكون أقل تأثراً بالعوامل الدخيلة مثل المستوى العالي من القدرة القرائية والتي تكون مطلوبة بالصيغ الأكثر تعقيداً.

3/ مثل هذه الفقرات تكون واضحة مفاهيمياً للمفحوصين حول طبيعة ما يسألون عنه أو المبدأ الأساسي الذي يختبرونه به.

ومن جهة أخرى اقترح فيريدريكسن (Fredriksen, 1981): من أجل فاعلية تصحيح أكبر أن مطوري الاختبارات يجب أن يركزوا كثيراً على الصيغ المركبة، والذي يجب أن يتوافر في

الأشياء التي تقاس بسهولة، وهذه قد تلقي ظلاً على ما يجب قياسه. وهذا قد يهمل اختبار الكثير من القدرات المهمة وبالتالي غير متعلمة. وأحد البدائل التي اقترحها فريديريكسن هنا هو تطوير مهام حل مشكلة غير مركبة متعددة البدائل. وتم وصف أعمال أكثر حداثة ذات علاقة بمثل هذه الاختبارات طرحها وارد وكارسون وويز تشلجر (Ward, Carson, Woisetsch-lager, 1983). وفي الوقت نفسه طرح قضية ما لو إذا كانت الفقرات التي تغطي المحتوى نفسه وباستعمال صيغ بديلة تقيس فعلاً السمة نفسها لم تحل بشكل حاسم ومقنع، ويبدو أن القضية لا زالت في تقديم مقترحات نظرية وأخرى امبريقية.

صياغة فقرات الاستبانات: Item formats for Inventories

هنالك صيغ ثلاث شائعة لاستبانات الشخصية والاتجاهات هي صيغة موافق – غير موافق، صيغة ليكرت، الصيغة الوصفية ثنائية الأقطاب، وفيما يأتي وصفاً لكل منها:

تتألف صيغة موافق– غير موافق على الأغلب من جملة تقريرية متبوعة باختيار واحدة من الاستجابتين. وهذه الصيغة موضحة بالفقرات الآتية:

1/ يجب أن يطيع الأطفال الوالدين دون أي سؤال	موافق	غير موافق
2/ يحتاج أطفال اليوم لضبط أقوى في البيت	موافق	غير موافق
3/ يثور معظم الأطفال الذين لديهم مشكلات سلوكية على الضبط الوالدي	موافق	غير موافق

والطريقة الأسهل لإعطاء الدرجات لمثل هذه الفقرات هو تقرير أي طرفي متصل الإتجاه سيُعطى الدرجة العالية وأيها سيعطى الدرجة المنخفضة. فمثلاً للفقرات المذكورة أعلاه، يمكن وصف الإتجاه على متصل يتراوح بين الإخضاع الكامل والتساهل من وجهة نظر والدية. وهنا يتوفر لمطور الاختبار فرصة لتخصيص المستجيب الأكثر سلطوية على أنه الذي يحصل على درجة أعلى على الأداء أو ان المستجيب الأكثر تساهلاً يحصل على الدرجة الأعلى. دعنا نفترض أن مطوِّر الاختبار يرغب أن تعكس الدرجات الأعلى للإتجاه الأكثر تساهلاً. بعدها يتم تحديد الدرجة التي تستحقها كل فقرة (إيجابية أم سلبية) بالنسبة للبناء. وفي مثالنا لأن المتصل معُرف بلغة التساهل فإن الفقرتين 1 ، 2 تُعد سلبية والفقرة 3 تعد إيجابية، وتصحيح الفقرات بإعطاء درجة واحدة لكل استجابة بالموافقة للفقرة الإيجابية. ونقطة واحدة لكل إستجابة بعدم الموافقة للفقرة السلبية، وعلامة المستجيب الكلية هي مجموع درجاته على الفقرات جميعها. وفي المثال الحالي المستجيب الذي أشر غير موافق، غير موافق، موافق

للفقرات (3,2,1) على التوالي فإنه سيحصل على الدرجات (1,1,1) ومجموعه 3 درجات. ومستجيب آخر أشر على غير موافق، موافق، غير موافق يكون نمط درجاته (0,0,1) ومجموعه درجة واحدة.

وطريقة بديلة لتصحيح فقرات صيغة موافق- غير موافق تستخدم أوزان الفقرات. فكل فقرة يؤشر لها بقيمة موزونة بحسب قوة التعبير الوجداني- الإنفعالي- نحو البناء الذي تقيسه، كذلك فإن هنالك طرائق عدة تستخدم في الحصول على أوزان الفقرات، والطريقة الأكثر شيوعاً هي طريقة الفترات المتساوية- القياس الفاصلي- الذي وصفه ثيرستون (Thur-stone, 1928). وعادةً يُعد كاتب الفقرات عدداً كبيراً من الفقرات (اقترح ثيرستون حوالي 100 فقرة) تتراوح بين موجب متطرف إلى سالب متطرف بالنسبة للبناء المقيس، بحيث يكون بعضها ذو تأثير متعادل. وقد تكتب كل فقرة على بطاقة منفصلة، وتعرض مجموعة الفقرات على مجموعة من المحكمين لتدريجها على متصل الفترات المتساوية، ويطلب من كل محكم قراءة الفقرة ووضعها على متصل مجزأ إلى 7 أو 9 أو 11 فترة متساوية في المسافات الفاصلة بينها. وعلى المتصل المؤلف من 11 فترة تقع الفقرة المتعادلة في الفئة أو الفترة السادسة.

وتحقق البيانات المستخلصة من عملية التحكيم هذه هدفين - الأول اختيار فقرات الصيغة النهائية للمقياس وتأشير الأوزان لكل منها، ووزن الفقرة هو وسيط تقديرات المحكمين. وعلاوة على ذلك تختار فقرات المقياس من المجموعة الكلية للفقرات على أساس إحصائي مثل Semi- interquartile range Q:

$$Q = \frac{م75 - م25}{2}$$

حيث ان م75 هو القيمة العددية للرتبة المئينية 75

- م25 هو القيمة العددية للرتبة المئينية 25 لمدى التوزيع.

وكلما كانت قيمة Q أقل كلما كانت الأحكام متفقة بدرجة أكبرفي قوة الإنفعال المعبر عنه بالفقرة، وهذه الفقرات هي الأفضل في تكوين المقياس. ومع ذلك فإن مطوِّر الاختبار يحاول تضمين المقياس فقرات من كل فئة تصنيف. وعندما يطبق المقياس على المستجيبين فيما بعد (وفي كل تطبيق) يُقر كل مستجيب فقرة تنطبق عليه أكثر من غيرها، ويضاف وزن هذه الفقرة إلى درجته الكلية.

والدرجة الكلية تقسم فيما بعد على عدد الفقرات التي أقرها المفحوص من أجل الحصول على متوسط وزن الفقرة. ويستخدم هذا المتوسط لحساب أو وصف إتجاهات المفحوصين أو المستجيبين.

والصيغة الثانية الأكثر شيوعاً للاستبانات اقترحها ليكرت (Likert, 1932)، وتتطلب طريقته كتابة مجموعة من الفقرات تبين كل منها إتجاه إيجابي أو سلبي بالنسبة للبناء المقيس. وهنا لا يتضمن المقياس فقرات محايدة، وعندما يقرأ المستجيب الجملة يختار استجابة على متصل مؤلف من خمس نقاط تتراوح بين موافق بشدة وغير موافق بشدة. وكما هو مبين على النحو الآتي:

يجب أن يطيع الأطفال والديهم بدون استفسار (دون معرفة الأسباب)

غير موافق بشدة   غير موافق   محايد   موافق   موافق بشدة

والفقرات هنا لا يتم تدريجها قبل تطبيقها على المستجيبين، ولكن يستخدم المستجيب متصل الإجابة (1-5) ليؤشر درجة إقراره للموقف المذكور في الفقرة. ولإعطاء الدرجات على متصل الاستجابة المتدرج تعطى الدرجة (1) للاستجابة التي تظهر أقل مستوى من التعبير الإنفعالي، والدرجة (2) للاستجابة التي تليها في درجة التعبير الإنفعالي، والدرجة (3) للاستجابة التي تليها في المستوى أو الشدة وهكذا. وإذا افترضنا ثانية أن مطور الاختبار يريد إعطاء درجات أعلى للمستجيب الأقل إنفعالاً أو أكثر تساهلاً، فإن الفقرات الثلاث المذكورة أعلاه تكون درجاتها على النحو الآتي:

الفقرة	موافق بشدة	موافق	محايد	غير موافق	غير موافق بشدة
1/ يجب أن يطيع الأطفال والديهم دون استفسار	1	2	3	4	5
2/ يحتاج أطفال اليوم ضبط أقوى في البيت	1	2	3	4	5
3/ يثور معظم الأطفال الذين لديهم مشكلات سلوكية على الضبط الوالدي	5	4	3	2	1

وعلامة المستجيب الكلية هي مجموع درجات الفقرات جميعها. وفيما يأتي بعض الإرشادات العامة التي تفيد في كتابة ومراجعة فقرات استبانة ليكرت أو موافق – غير موافق:

1/ اكتب الجمل أو الفقرات في صيغة الحاضر.

2/ لا تستخدم جمل واقعية أو تلك التي تكون قابلة للتفسير الواقعي.

3/ تجنب الجمل التي تحتمل أكثر من تفسير.

4/ تجنب الجمل التي يجيب عنها الجميع بـ غالباً (أو موافق بشدة) أو نادراً (أو غير موافق بشدة).

5/ حاول أن تكون عدد الفقرات التي تعبر عن انفعال إيجابي مساوياً لعدد الفقرات التي تعبر عن انفعال سلبي.

6/ أن تكون الجمل قصيرة ولا تتجاوز 20 كلمة.

7/ أن تكون الجمل صحيحة من حيث قواعد اللغة.

8/ يجب تجنب الجمل التي تتضمن كليات مثل لا شيء، وأبداً لأنها تعطي الغموض.

9/ تجنب استخدام متطلبات غير محددة (indefinite qualifiers) مثل فقط، دائماً، ليس غير، عديد، قليل، نادراً.

10/ كلما أمكن، حاول أن تكون الفقرات بسيطة، وليست جمل معقدة أو مركبة، وتجنب الجمل الشرطية التي تستخدم (إذا) أو (لأن).

11/ استخدم الألفاظ التي تفهم بسهولة من قبل المستجيبين.

12/ تجنب استخدام علامات النفي مثل (لا، ولا واحدة، أبداً).

ويمكن الرجوع إلى مصادر حديثة تفيد في بناء فقرات مقاييس الإتجاه مثل اندرسون (Anderson, 1981) وداويس (Dawes, 1972) ويودنسكي وزملاؤه (Undinsky, et.al., 1981).

وهناك صيغة أخرى شائعة في بناء الاستبانات هي استخدام أزواج الصفات القطبية. ويعود أصل هذه الصيغة إلى أوزجود وزملاؤه (Osgood, et.al., 1957) إذ اقترحوا استخدامها في دراسة المعاني اللغوية للأبنية النفسية، إذ يكتب في الأعلى اسم البناء النفسي المراد قياسه ويتبع بزوج من الصفات التي تمثل قطبين متعاكسين على المتصل نفسه، ويفصل بينهما إما (5 أو 7) نقاط على المتصل، ويوجه المفحوص ليضع علامة على المتصل في المكان الذي يعكس مشاعره. وكمثال في محاولة تقويم إتجاهات المعلمين في العمل مع الطلبة المعاقين عقلياً، يمكن استخدام الصيغة الآتية:

جميل ———————— قبيح

سعيد ———————— حزين

قذر ———————— نظيف

وبعد أن اختبر أوزجود وزملاؤه عدداً كبيراً من أزواج الصفات المتنوعة وجدوا أن معظمها يتجمع ضمن واحدٍ من الأبعاد الثلاثة الآتية: التقويم ، والفعالية، والنشاط والتي تقع ضمن المعنى اللغوي للبناء اللفظي. وقد اقترحوا طريقة في تحليل استجابات هذه الفقرات وتفسيرها، وأطلقوا على أداتهم هذه اسم التفاضل اللغوي. وعلى مطور الاختبار الذي يعد مجموعة متميزة من الصفات ثنائية القطب الجديدة (أي دون استخدام الأزواج المختبرة من قبل أن يحدد كيفية إعطاء الدرجات وتفسيرها. وفي مثل هذه المواقف يجب وصف الفقرات في قائمة صفات ثنائية القطب بدلاً من مقياس التفاضل اللغوي.

ومن المهم عند تطوير فقرات المقياس الإنتباه إلى مجموعة الاستجابات التي تؤثر على سلوك المفحوصين منفردين. ويمكن تحديد مجموعة الاستجابات على أنها ميل المفحوص لأن يستجيب بطريقة معينة لصيغة معينة من الفقرات بغض النظر عن محتواها.

وقد قدم جيلفورد **(Guilford, 1954)** تحديداً تقليدياً لأنواع عديدة من الاستجابات المألوفة، وقدم مقترحات لخفض أو ضبط أثر هذه التأثيرات. ومجموعتين من الاستجابات تؤثر في الدرجة على المقياس هي الميل للموافقة على جملة بغض النظر عن محتواها والتفسيرات التفاضلية للأفراد على الأشياء والمتطلبات غير المحددة مثل بعض، عادة.

وهناك قضايا أخرى تتعلق ببناء الاستبانات تتضمن أثر استخدام أنواع مختلفة من النقاط الرابطة (anchor points) على متصل الإجابة انظر (Frisbie & Branden burg, 1979) و (Lam & Klockers, 1982)، أو عدد مختلف من بدائل الاستجابة (أنظر Masters, 1974 و Velicer & Stevenson, 1978). ومثل هذه الدراسات توضح أن القرارات حول كيفية قياس السلوك يمكن أنْ تثبّت بصورة حتمية الشيء المقيس.

مراجعة الفقرات Item Review

عندما تعد مسودة الفقرات ينصح مطور الاختبار بأن يراجع زملائه المختصين وبشكل غير رسمي مراجعة الفقرات لكل من الدقة، والصياغة، والقواعد، والغموض وأمور فنية أخرى، وبعدها يمكن مراجعة الفقرات التي تتضمن مشاكل عند الضرورة. إضافة إلى ذلك وحالما تكتب الفقرات يجب أن تخضع لمراجعة رسمية فقرة فقرة (Item by item)، ويجب أن يتضمن

ببناء الفقرة الآخذ بعين الاعتبار الأمور المهمة الآتية:

1/ الدقة

2/ الملائمة أو المناسبة لمواصفات الاختبار

3/ الخلل الفني في بناء الفقرة

4/ القواعد

5/ الهجومية أو التحيز

6/ مستوى المقروئية

وهذا يتطلب خبراء متنوعين في مراجعة الفقرات، فمثلاً، الخبراء في الموضوع يكونون الأكثر ملائمة لتقرير، ما إذا كانت الفقرات واضحة ومتضمنة المفاتيح الصحيحة (Correct-ly Keyed)6، وهؤلاء الخبراء هم المؤهلون للحكم على ما إذا كانت الفقرات مناسبة لمواصفات الاختبار أو خصائص الاختبار (وتفصيلات أكثر لهذه العملية مبين في الفصل العاشر ضمن موضوع صدق المحتوى). كذلك فإن المختصين في القياس وبناء الاختبارات هم الأنسب للحكم على ما إذا كان هنالك خللاً في البناء المقيس أولاً، فمثلاً عندما تكون الفقرات من نوع الاختيار من متعدد فإن الخبير الفني ينظر إلى الخلل في الصيغة كأن يكون البديل الصحيح أطول من البدائل الأخرى، كذلك هو الأقدر على تحديد أنماط الاستجابة التي يبدو أنها تؤثر في الدرجات، وهو يأخذ بعين الاعتبار ملائمة الفقرات لأنماط الاستجابة المختلفة. ومن المهم توافر خبير لغوي إذ أن الفقرات يجب أن تخلو من الأخطاء النحوية بما فيها من أخطاء تهجئة وأخطاء تنقيط أو استخدام جمل وعبارات غير مألوفة مما يؤدي إلى تفسير خاطيء للفقرة. من جهة أخرى على الخبراء جميعاً الإنتباه إلى الأخطاء واللغوية.

ومن الضروري أن يعرف واحدٌ أو أكثر من فريق الخبراء خصائص المجتمع الذين ستطبق عليهم الفقرات، من أجل أخذ الحيطة والحذر من كون الفقرات تأخذ طابع عدواني أو متحيز ضد أي مجموعة فرعية من المجتمع الكلي، مثل استعمال مظهر حضاري غير مألوف أو محتوى غير مألوف للمجموعات الفرعية، وذلك عندما يكون المحتوى مرتبطاً بالبناء المقيس ارتباطاً مباشراً (على سبيل المثال: في مسألة رياضية تتطلب حساب الربح البسيط «يجب أن تستخدم غرضاً ملائماً للشراء من قبل معظم المفحوصين أو عائلاتهم لا استخدام غرض شرائه محصور بالفئة المترفة أو العائلات المقتدرة فقط».

وعندما تعد الفقرات للأطفال أو المراهقين يجب أن يؤخذ بعين الاعتبار مستوى مقروئية الفقرات عندما يكون هدف الاختبار قياس شيء آخر غير مهارات القراءة. وهناك أساليب

معيارية لتقييم مقروئية مقالة مكتوبة التي لا تناسب فقرات الاختبار بسبب تركيبها أو طولها، وهنالك طريقة واعدة لقياس المقروئية قدمها ايرونسون وزملاؤه (Ironson, etal., 1984).

هذا، ويمكن القيام بمراجعة الفقرات قبل التطبيق الأولي أو بعده، فاختيار التسلسل المناسب يكون على أساس من الملائمة والاقتصاد. فإذا توفر الخبراء المراجعون ووقتهم يسمح بالمراجعة فإن المراجعة تكون قبل التجريب الأولي، وبالتالي فإن وقت التجريب لا يضيع في اكتشاف الفقرات الخاطئة أو المتحيزة. ومن جهة أخرى تحتاج العديد من الفقرات مراجعة أو إعادة صياغة بعد التجريب، مما يؤدي إلى لزوم مراجعة إضافية للفقرات من قبل الخبراء. أيضاً إن كان الوصول للخبراء ثانية مكلفاً من حيث الوقت والجهد فإن العديد من معدي الاختبارات ومطوريها يؤجل أحكام الخبراء إلى ما بعد التجريب الأولي للفقرات وبعد المراجعات المستفيضة. وعندما تستخدم نتائج تحكيم الخبراء على أنها مؤشرات وشواهد لصدق المحتوى فمن المهم أن تكون المراجعة من قبل فريق المحكمين للصيغة النهائية للاختبار أو الفقرات.

التجريب الأولي للفقرات: Preliminary item tryouts

من المستحسن لمعد الاختبار قبل أن يطبع الفقرات بصيغتها النهائية تجريب الفقرات على عينة صغيرة من المفحوصين، وعندما يتوافر عدد قليل من الأفراد (كما هو الحال في البرامج التجريبية)، فإنه يجب الاحتفاظ بمعظم المفحوصين وذلك لتجريب الفقرات ميدانياً على هؤلاء. وفي مثل هذه الحالات من الضروري تطبيق الفقرات على عدد قليل من الأفراد يتراوح عددهم بين 15-20. بينما تطبق الفقرات على عينات أكبر يتراوح عددها بين 200-300 مفحوصاً، وذلك عندما يطوّر الاختبار للاستخدام التجاري، وعندما يكون عدد الفقرات كبيراً جداً وفترة التطبيق قصيرة يمكن تطبيق مجموعات من الفقرات على عينات مختلفة من المفحوصين. ويكون التجريب الأولي للفقرات على الأغلب غير رسمي، وعلى مطور الاختبار أن يستغل هذه الفرصة ليلاحظ تفاعل المفحوصين مع الفقرات وبعض سلوكاتهم مثل التوقف الطويل أو الخربشة أو تبديل الإجابة، وهذه تؤشر تشويش في فقرات بعينها. وبعد التجريب يجب عمل مراجعة تستخلص منها ملاحظات حول الفقرات جميعها، ومن ثم تقديم مقترحات لإجراء التحسينات الممكنة.

ويوصى أيضاً بإجراء تحليل احصائي وصفي للاستجابات على كل فقرة، وهذه تمكن مطور الاختبار من الحصول على فكرة سريعة عما إذا كانت الفقرات ملائمة من حيث مستوى الصعوبة للمجموعة ككل، وما إذا كان هنالك تبايناً كافياً في الاستجابات من أجل تعديل ما

سبق ذكره ليلائم الاختبار عند تطبيقه على مدى واسع، كذلك يجب مراجعة الفقرات بشكل مكثف بالإفادة من نتائج التجريب الأولي.

الخطوة التالية The next step

بعد التجريب الأولي والدراسات المستفيضة المتتابعة للفقرات تكون جاهزة للتطبيق على نطاق واسع (التطبيق الميداني)، وهذا يتطلب تطبيق الفقرات بصيغتها النهائية على عينة كبيرة من المفحوصين تمثل المجتمع الذي صمم الاختبار من أجله. كذلك دراسة الخصائص الاحصائية لدرجات الاختبار من خلال طرائق متنوعة تعرف باسم تحليل الفقرات.

وفي الوحدة السابعة عشر من هذا الكتاب تعرض دراسات تصميم الاختبار الميداني للفقرات وإجراء التحليل المناسب. وعندما يكون التفسير معياري المرجع هو الهدف فإن مطور الاختبار يستخدم نتائج التحليل في طرح أو حذف الفقرات التي لا تؤدي الوظيفة كما يجب. وكما سنتعلم من الفصل الرابع عشر إنه لا يوجد اتفاق على استخدام نتائج التحليل في حذف فقرات من الاختبارات محكية المرجع.

وعندما يتم إعداد الصيغة النهائية للاختبار، فإن مطور الاختبار ملزم بإجراء دراسات ثبات وصدق درجات الاختبار. النظريات والتطبيقات المناسبة للثبات هي موضوعات الفصول من السادس وحتى التاسع (الوحدة الثانية من الكتاب). والمقترحات لتطوير طرائق التصحيح، واستخراج المعايير وتقديم البيانات المعيارية للإفادة منها في تفسير درجات الاختبار هي موضوع الوحدة الخامسة.

الخلاصة:

تم في هذا الفصل تحديد عشر خطوات منتظمة متسلسلة في تطوير الاختبار. الأول: يجب تحديد الهدف أو الأهداف التي تستخدم فيها درجات الاختبار بوضوح، ويجب اختيار الأسلوب الأكثر ملائمة في تطوير الاختبار في ضوء هذه الأهداف. وفي الثاني تم وصف أسلوبين ممكنين في تطوير الاختيار، وهما تطوير اختبار يميز بين الأفراد على بناء معين، وتطوير اختبار مصمم لإنتاج درجات تصف مستويات كفاءة أكثر تجريد في مجال محتوى معين. وخطوات مهمة في تحديد السلوكات المتعلقة ببناء معين تتضمن تحليل المحتوى ومراجعة الأبحاث والمواقف المهمة والملاحظات المباشرة والأهداف التعليمية. وفي تطوير الاختبارات التحصيلية المحكية فإن هدف مطور الاختبار يكون تأسيس نطاق

فقرات محدد جيداً بحيث أن مستخدمي الاختبار يمكنهم الاستدلال من درجات اختبار مؤلف من عينة فقرات على الكفاءة لنطاق الفقرات الكلي. وكنتيجة فإنه يمكن تأسيس نطاق الفقرات من خلال الأهداف التدريسية وخصائص الفقرات وصيغ الفقرات.

والخطوة الثالثة في بناء الاختبار هو تطوير جدول المواصفات الذي يبين نسب الفقرات في الاختبار التي تركز على فئات المحتوى والعمليات المتنوعة التي تناسب البناء الذي نهتم به (أو نطاق الفقرات). وتم عرض عدة تصنيفات لفئات السلوك في المجال المعرفي، وأكثرها شهرة هو ما اقترحه بلوم (1956)، إضافة إلى التصنيف المقترح في المجال الوجداني الذي وصفه كراثول وبلوم وماسيا (1964).

وفي الخطوة الرابعة يتم بناء الفقرات التي تتضمن اختيار الصيغة الملائمة للفقرات وإثبات أن هذه الصيغة واضحة تحدد الأهداف المنشودة لمجتمع المفحوصين المستهدفين في الاختبار- وبعدها اختيار كتبة الفقرات وتدريبهم وتوجيه عملية كتابة الفقرات. وهنالك ثلاث صيغ شائعة في اختبارات أقصى أداء هي اختيار البديل، والمزاوجة، والاختيار من متعدد. وهنالك ثلاث صيغ واسعة الإنتشار للاستبانات هي صيغة موافق- غير موافق المستخدمة في تدريج ثيرستون، وصيغة ليكرت، وقائمة الصفات القطبية (ثنائية القطب)، وقد تم وصف صيغ هذه الفقرات جميعها باختصار.

والخطوة الخامسة هي مراجعة الفقرات من قبل فريق من الخبراء الذين يجب أن يأخذوا بعين الاعتبار بعض الأمور مثل الدقة والملائمة مع مواصفات الاختبار والنوعية والقواعد وأثر العامل الثقافي أو التحيز لجنس والمقروئية. إضافة إلى أنه يجب تطبيق الفقرات على عينة صغيرة من المفحوصين وذلك للتحقق من أن تطبيق المفحوصين للتعليمات الخاصة بالإختبار، وتقدير الوقت اللازم للاختبار، ولتحديد الفقرات الغامضة أو التي منها قصور في الكتابة. ويجب تنفيذ المراجعات التي يؤشر إليها من خلال مراجعة الفقرات أو التطبيق على العينة الاستطلاعية قبل تطبيق الاختبار ميدانياً على عينة كبيرة. والفقرات الأخيرة في

تطوير الاختبار تتضمن التطبيق الميداني وتحليل الفقرات ودراسات الصدق والثبات ووضع إرشادات لتفسير درجات الاختبار. وتم تغطية هذه الطرائق في الفصول التالية من الكتاب.

التمارين:

1/ راجع الأهداف من أ إلى ل بالاعتماد على محتوى الفصل الثاني، وصنف هذه الأهداف في خلايا جدول المواصفات التي يليها:

أ- تفسير خصائص توزيع درجات مجموعة مفحوصين من خلال مقاييس النزعة المركزية ومقاييس التشتت.

ب- حساب وتفسير المقاييس الشائعة للنزعة المركزية والتشتت.

جـ- تمييز معاني المفاهيم مثل: المتغير المتقطع، المتغير المتصل، التكرار، التكرار التراكمي، الاحتمال التراكمي، مقاييس النزعة المركزية، مقاييس التشتت، القيمة المتوقعة، المنحنى الطبيعي، المنحنى المعياري الطبيعي، الدرجة الخام، الدرجة المنحرفة، الدرجة الزائية، معامل الارتباط، الأزواج المرتبة، معامل الانحدار، معادلة التنبؤ، الخطأ المعياري للتقدير.

د- تحويل الدرجات الخام إلى الدرجات المنحرفة والدرجات الزائية.

و- ربط الدرجات الزائية المعتمدة على المنحنى الطبيعي بالمساحة تحت المنحنى الطبيعي باستخدام الجداول الزائية الطبيعية.

و- تمييز المواقف التي تكون فيها الدرجات الزائية أكثر فائدة من الدرجات الخام.

ز- استخدام الأزواج المرتبة لعمل تفسيرات حول طبيعة العلاقة بين المتغيرات.

ح- حساب معامل الارتباط وتفسيرها.

ط- تفسير العلاقة بين الأزواج المرتبة وخط انحدارها.

ي- حساب قيم الميل والقاطع لخط الانحدار لبيانات واستخدامها في بناء خط الانحدار.

ك- استخدام معادلة الانحدار في حساب القيم المُتنبأ بها للأفراد على متغير محكي بمعرفة درجته على المتنبيء.

ل- استخدام الخطأ المعياري للتقدير في تكوين فترة ثقة حول قيمة درجة المحك المتنبأ بها.

	توزيع المعالم	التوزيع الطبيعي	الارتباط / الانحدار	
المعرفة				%10
الفهم				%30
التطبيق				%60
	%35	%15	%50	

2/ افترض أنك تريد بناء اختبار يغطي وحدة الاحصاء باستخدام جدول المواصفات السابق ولديك الوقت الكافي لبناء 30 فقرة.

أ- ما نسبة الفقرات التي يجب كتابتها بمستوى المعرفة، والفهم، والتطبيق على التوالي.

ب- ما نسبة الفقرات التي يجب أن تغطي موضوعات مواصفات التوزيع الأساسية والارتباط على التوالي.

جـ- ما نسبة الفقرات التي يجب أن تغطي موضوعات الارتباط والانحدار في مستوى التطبيق.

3) راجع مواصفات الفقرة المبينة في شكل (4-1)، وطوِّر مواصفات مشابهة للمهارة الفرعية «القدرة على طرح الأعداد العشرية».

4/ أ) أنظر إلى المشكلة الآتية المطروحة في اختبار فيزياء للمرحلة الثانوية «جمعت كمية من الغاز فوق الماء، عند درجة 16°م، وكان ضغط خليط الغازات 982.9 تور. وسحب تور بخار الماء، وكان الضغط الجزئي للغاز المتبقي 969.3 تور. فما ضغط الماء عند درجة 16°م. إذا أراد المعلم أن يحوِّل هذه المسألة إلى صيغة سؤال. حدد جزئي السؤال التي يجب أن تضمم على أنها موقع مجموعات الإحلال.

ب) افترض أن المعلم صمم هذه الفقرة بالأساس على أنها تقيس عملية التطبيق. وطبق

الاختبار على صفوف عدة من الطلبة الذين تقدموا للاختبار على فترات متتالية في اليوم الدراسي. ولأن المعلم خاف من أن الإجابات أظهرت صيغة مختلفة للاختبار مع كل صف، من خلال صيغ الفقرة والتعويض بقيم مختلفة في المسألة. ولاحظ المعلم فيما بعد أن الصفوف التي تقدمت للاختبار في آخر اليوم كان أداؤها أفضل من تلك التي تقدمت للاختبار في بداية اليوم، إلى ماذا تُعزِي هذه النتيجة.

جـ) قارن بين مزايا ومساوىء استخدام أسلوب الفقرة - الصيغ في بناء الاختبار للموقف السابق.

هـ) راجع فقرات استبانة الاتجاهات وحدد القصور الفني الذي يجب تصحيحه.

أ- يجب أن تبقى الأمهات العاملات في البيت لأنهنّ يأخذن الوظائف التي يحتاجها الرجال المسؤولون عن عائلات.

ب- إذا أرادت الأم أن تعمل، فيجب أن يكون في الوقت الذي يكون فيه أطفالها في المدرسة.

جـ- في الولايات المتحدة، أكثر من نصف الأمهات اللواتي أطفالهن في عمر المدرسة يعملن خارج البيت.

د- ليس من الممكن للمرأة أن تجمع بين رغبتها في الحصول على وظيفة مع عدم تقصيرها بوظيفتها الأساسية كأم دون أن تشعر بالضغط.

6/ طالب دراسات عليا في علم النفس التطوري مهتم في بناء استبانة لتقييم مفهوم الذات للأطفال. اقترح على الأقل ثلاث طرائق تفيده في تحديد النطاق لهذا البناء.

7/ عالم نفس مهتم في تطوير استبانة لتقييم اتجاهات المستخدمين في المجتمع نحو العاملين في مجال خدمة المعوقين حركياً.

أ- حدد ثلاث طرائق مختلفة يمكن استخدامها في تحديد البناء، وقارن بين مزايا ومساوىء استخدام كل منها.

ب- حدد ثلاث صيغ لفقرات مختلفة يمكن استخدامها في هذا الموقف، واكتب فقرة واحدة تمثل كل منها.

8/ للمناقشة الصفية، إقرأ مقالة (Irvin, Halpern & Landman, 1980). أجب عن الأسئلة الآتية وناقش إجاباتك.

أ- هل تبين المؤلفون (أو أشاروا بوضوح) إلى الهدف الأساسي التي تستهدفها درجات الاختبار في بطارية المعلومات الاجتماعية وما قبل المهنية.

ب- أي أساليب تطوير الاختبار الأساسية يبدو أنها اعتمدت في عملية البناء- الفروق الفردية أن معاينة النطاق.

جـ- قارن وصف المؤلفين للدراسة مع الخطوات العشرة المذكورة عن تطوير الاختبار في هذا الفصل. وأي من هذه الخطوات حذفت؟ حدد الأنشطة التي وصفها المؤلفون وتناظر خطوات معينة متبعة في تطوير الاختبار ووضعت في هذا الفصل.

د- إلى أي مدى تدعم الارتباطات الناتجة فرضية أن فقرات اختيار البديل وفقرات الصح ... الخطأ وفقرات الاختيار من متعدد تقيس بناءات مختلفة.

الفصل الخامس

5

درجات الاختبار كدرجات مركبة

الفصل الخامس

درجات الاختبار كدرجات مركبة

تعد بطارية الاختبارات تجميعاً لاختبارين أو أكثر صممت لتطبق على مجموعة المفحوصين نفسهم، وتحسب فيها درجة منفصلة لكل مفحوص على كل اختبار في البطارية. ومن الأمثلة المشهورة لهذه الاختبارات اختبار GRE، الذي يمكن بوساطته الحصول على درجات منفصلة على اختبارات لفظية، وكمية، وتحليلية لكل مفحوص. والدرجة المركبة هي الدرجة الناتجة عن جمع درجات اختبارين فرعيين أو أكثر، وللتفسير الصحيح للدرجات المركبة من المهم فهم كيفية تأثير خصائص درجات الاختبارات الفرعية احصائياً على الدرجة المركبة. والأكثر أهمية من ذلك هو أن يميز باني الاختبار ومطوره أن درجة كل اختبار تعد درجة مركبة، ويمكن تحديد درجة الفقرة على أنها عدد النقاط المُعطاة لاستجابة المفحوص على فقرة معينة. وعند إعطاء درجات الاختبار فإن درجة الاختبار الكلية تحدد عادةً بجمع درجات الفقرات جميعها. لهذا فإنه يمكن اعتبار كل فقرة في الاختبار على أنها اختبار فرعي قصير. ويمكن اعتبار درجات الفقرات على أنها درجات اختبارات قصيرة (درجات اختبار صغرى). ومن هذا المنطلق فما أن يتم حساب درجة الاختبار بجمع درجات فقراته فإن الدرجة الكلية للاختبار تعد درجة مركبة.

وسيطرح هذا الفصل الإجراءات الاحصائية الممكنة لدرجات الفقرة وعلاقة خصائصها بدرجة الاختبار الكلية. إن الشيء الأساسي والمهم في عملية تطوير الاختبار هو الفهم بأن هذه الفقرات تؤلف الوحدات البنائية للاختبار. ولا يكون للاختبار ككل خاصية غير تلك الخصائص الوظيفية لفقرات الاختبار. ولكل من ينشد بناء اختبار جيد يجب أن تتوافر لديه معلومات حول توزيع درجات الفقرات وعلاقة كل فقرة بالفقرات الأخرى المؤلفة للاختبار.

وقد بحثنا في الفصل الثاني كيفية تطبيق مفاهيم الوسط والتباين والانحراف المعياري في وصف درجات الاختبار. وسنعرض الآن كيفية تطبيق هذه الاحصائيات على درجات الفقرات ومن ثم كيفية تأثيرها على توزيع خصائص الدرجة الكلية.

مخططات تصحيح الفقرة

يمكن تصنيف فقرات الاختبار على أنها ثنائية التفرع أو متعددة التفرع. وتعد الفقرة ثنائية

التفرع إذا كانت القيم الوحيدة الممكنة للدرجة إما صفراً (إجابة خاطئة) أو واحداً (إجابة صحيحة). وفيما يأتي أمثلة لفقرات ثنائية التفرع.

1- تسمى المسافة بين نهايات الخلايا العصبية:

أ- شجيرات عصبية ب- محور عصبي

جـ نقطة التشابك د- نيوترون

2- يجب أن يكون للمعلمين في المدارس الحق في الاضراب:

أ- موافق ب- غير موافق

ففي الفقرة الأولى تعطى الإجابات (أ، ب، جـ) الدرجة صفر والإجابة (جـ) الدرجة (1). وفي الفقرة الثانية تعطى الإجابة (أ: موافق). الدرجة (1) والإجابة (ب: غير موافق) الدرجة صفر.

وكما رأينا في المثالين السابقين فإن الفقرات ثنائية التفرع قد تظهر في مقاييس كلا المجالين المعرفي والوجداني، فقد تكون ذات بدائل متعددة كما في المثال الأول، وقد تكون ذات بديلين كما في المثال الثاني بما أن التصحيح يعطي الدرجة صفر أو واحد فقط. أما مخططات التصحيح غير الثنائية فإن مدى الدرجات الممكنة يكون غير مقيد بواحد أو صفر. وكأمثلة نذكر:

1/ أكتب جملة صحيحة قواعدياً باللغة الألمانية مستخدماً الصيغة المفردة للفعل Verstehen (أعلى درجة للفقرة 3، ويمكن إعطاء أي جزء منها).

2/ يعد الشخص المتخلف عقلياً غير منتج في المجتمع:

أ- موافق بشدة ب- موافق جـ- لا رأي لي

د- غير موافق هـ- غير موافق بشدة.

(وتتدرج درجات الفقرة من 1 إلى 5 مع أعلى درجة للإتجاه الإيجابي للمواطنين نحو الأفراد المتخلفين عقلياً).

وكما هو مبين في المثال، فإن الفقرات المقالية أو ذات الإجابة القصيرة تصحح على الأغلب باستخدام مخططات غير ثنائية التفرع، وكذلك فقرات معظم مقاييس الاتجاهات أو الأداء فإنها تصحح بهذه الطريقة.

الفقرات غير ثنائية التفرع تخضع الصيغ التقليدية للوسط والتباين
استجابات تباين الأفراد — والارتباط في وصف توزيع
الوسط يستخدم لتحديد قوة اتجاه الفقرة — درجات الفقرات

الاحصاء الوصفي للدرجات غير ثنائية التفرع

الارتباط بارتباط الفقرات مع بعضها البعض

يبين الجدول (5-1) استجابات (10) مفحوصين على (5) فقرات في مقياس اتجاهات تم تصحيحه على تدريج يمتد من «أوافق بشدة» إلى «لا أوافق بشدة». وتتراوح القيم الممكنة لدرجات كل فقرة (من 1 إلى 5)، ويطلق على مثل هذا الشكل اسم مصفوفة الفرد- الفقرة. ويتضمن كل صف في المصفوفة استجابات الفرد على فقرات الاختبار جميعها، ويتضمن كل عمود استجابات الأفراد جميعاً على فقرة بعينها.

جدول: (5-1): استجابات 10 أفراد على 5 فقرات في مقياس اتجاهات صححت على تدريج من 1 إلى 5:

الفرد	الفقرات 1	2	3	4	5	المجموع
1	5	5	4	3	2	19
2	2	2	3	1	2	10
3	4	4	3	3	2	16
4	2	2	2	1	2	9
5	5	5	3	5	4	22
6	1	1	2	2	3	9
7	1	2	3	1	1	8
8	4	1	3	4	5	17
9	5	3	4	4	3	19
10	2	2	3	3	4	14
وسط الفقرة	3.1	2.7	3.0	2.7	2.8	
تباين الفقرة	2.5	2.0	0.4	1.8	1.4	

وفي هذه المصفوفة يشار إلى موقع كل عنصر برمز كبير بذيله رمزين فرعيين (Xrc)، إذ يمثل الرمز الفرعي الأول (r) قيمة الصف، ويمثل الرمز الفرعي الثاني قيمة العمود. وهكذا فإن (X_{11}) تشير إلى الدرجة في الصف الأول في العمود الأول أو استجابة الفرد الأول على الفقرة الأولى. وفي الجدول أعلاه فإن قيمة (X_{11})، تساوي (5). وأما الرمز (X_{23}) فإنه يشير إلى قيمة الصف الثاني في العمود الثالث، أو استجابة الفرد الثاني على الفقرة الثالثة وتساوي (3). وإن كان معد الاختبار مهتماً بتحديد قوة الاتجاه لفقرة معينة فما عليه إلا أن

يحسب وسط الدرجات على الفقرة. ونرمز إلى وسط الدرجة i بالرمز μ_i، وتحسب باستخدام المعادلة (2-1)، وكما يأتي:

$$\frac{\sum X_1}{N} = \mu_1$$

$$10/2+5+4+1+1+5+2+4+4+5 =$$

$$.3.1=$$

وإنْ كان يريد معرفة مدى تباين استجابات الأفراد على فقرة، فإن تباين الفقرة يحسب بتطبيق المعادلة (4.2). وعلى النحو الآتي للفقرة الأولى:

$$\frac{\sum(X-\mu 1)^2}{N} = \sigma_1^2$$

$$10/\ ^2(3.1\ -2) +\ 000\ +\ ^2(\ 3.1 - 2) +\ ^2(3.1 - 5) = \sigma_1^2$$

$$2.5 =$$

وقد يهتم معد الاختبار أيضاً بمعرفة ما إذا كان هنالك علاقة بين درجات الأفراد على الفقرتين (2,1)، وهذه تحسب باستخدام المعادلة (2-13) والتي تبين كيفية حساب معامل الارتباط. ويبين الجدول (5-2) مصفوفة معاملات الارتباط لكل زوج من الفقرات الخمس المذكورة في الجدول (5-1):

جدول (5-2): مصفوفة الارتباط الداخلية لدرجات الفقرات المدونة في جدول (5-1):

الفقرة	1	2	3	4	5
1	1.00	.73	.70	.81	.34
2		1.00	.56	.48	.16-
3			1.00	.47	.00
4				1.00	.73
5					1.00

ويتضح لنا من الجدول أن الفقرات (4,3,2,1) تؤشر أنماط استجابة متشابهة، بينما الفقرة (5) ترتبط بالفقرة (4)، ولكن الأداء على هذه الفقرة له علاقة ضعيفة بالاداء على

الفقرتين (3,2). وسنناقش في الأجزاء التالية كيف أن معلومات مثل هذه تفيدنا في بناء الاختبار وتنقيحه.

الإحصاء الوصفي للفقرات ثنائية الدرجة

الوسط والتباين

عندما تكون الفقرات ثنائية التفرع فإنه يمكن استخدام صيغة مبسطة لحساب وسط الفقرة وتباينها وانحرافها المعياري. ولاشتقاق هذه الصيغ سنستخدم مفهوم يعرف بـ: «صعوبة الفقرة»، والتي تعرف بأنها نسبة الأفراد الذين أجابوا على الفقرة إجابة صحيحة. وعلينا ملاحظة أن هذا المصطلح تقني ويبدو متناقضاً ومثيراً للجدول عند استخدامه من قبل الأفراد العاديين. ووفقاً لتعريف صعوبة الفقرة فإن الفقرة التي أجاب عليها (85%) من أفراد العينة إجابة صحيحة فإن صعوبتها (أو قيمة P لها) تساوي (.85)، بينما الفقرة التي أجاب عليها (50%) من أفراد العينة إجابة صحيحة فإن صعوبتها تساوي (.50)، أي بنسبة صعوبة أقل. وهكذا فإن الفقرة الأسهل يكون معامل صعوبتها أعلى. ويبين الجدول (5-3) مصفوفة الفرد-الفقرة التي تمثل كل عنصر فيها درجة الفرد (i) على الفقرة (j). وفي أسفل الجدول معامل صعوبة كل فقرة. وهذا ليس صعباً لأن:

$$\text{صعوبة الفقرة} = \frac{\text{عدد الأفراد الذين حصلوا على درجة على الفقرة}}{\text{عدد أفراد العينة}}$$

دعنا الآن نتذكر أن وسط الفقرة يحسب باستخدام الصيغة:

$$\mu_j = \frac{\sum x j}{N}$$

وحيث أن قيم (x) إما (1 أو صفر) فإن $\sum xj$

تكافيء عدد الأفراد الذين حصلوا على الدرجة (1) على الفقرة.

وهكذا فإن:

$$P j = \mu j \qquad (1\text{-}5)$$

جدول (3-5): استجابات 10 مفحوصين على (5) فقرات صححت ثنائياً:

المفحوصين	الفقرات					المجموع
	1	2	3	4	5	
1	0	0	1	1	0	2
2	0	0	0	1	0	1
3	1	1	0	0	0	2
4	1	0	0	1	0	2
5	0	1	1	1	1	4
6	0	1	0	0	0	1
7	1	1	1	1	1	5
8	1	1	0	1	0	3
9	1	1	1	1	0	4
10	0	0	0	1	1	2
Pi (وسط الفقرة)	0.50	0.60	0.40	0.80	0.30	
Pjqi (تباين الفقرة)	0.25	0.24	0.24	0.16	0.21	

كذلك، يمكن حساب تباين الفقرة باستخدام صيغة التباين الشائعة:

$$\sigma_i^2 = \frac{\sum (X_j - \mu)^2}{N} \qquad 2\text{–}5$$

وللفقرات ثنائية التصحيح، فإن الصيغة الأبسط لحساب تباين الفقرة هي:

$$\sigma_i^2 = P_i\, q_i \qquad 3\text{-}5$$

حيث أ ب $q_i = (1\text{-}Pi)$. ويمكن عرض أسلوب جبري بسيط لإثبات الصيغة (3-5) لتبيان أن الصيغة (3-5) مشتقة مباشرة من الصيغة (2-5). إذ يمكن فك القوس في المعادلة (2-5) على النحو الآتي:

$$\sigma_i^2 = \frac{\sum X_{ij}^2}{N} - \frac{2\sum X_{ij} - \mu_j}{N} - \frac{\sum \mu_i^2}{N} \qquad 4\text{-}5$$

ولأن X_{ij} تساوي إما (1 أو صفراً) فإن:

$$\frac{\sum X_{ij}}{N} = \frac{\sum X_{ij}^2}{N}$$

وهذه تساوي P_i، كذلك فإن μ_j هي قيمة ثابتة لقيم X_{ij} جميعها، فإن الحد الثاني في المعادلة يمكن كتابته على النحو الآتي:

$$\frac{-2\sum X_{ij}\ \mu_j}{N} = -2\mu_j^2 = 2P_j^2$$

كذلك فإن الحد الأخير في المعادلة (5-4): $\frac{\sum \mu 2_j}{N}$ له قيمة

ثابتة أيضاً مضاعفة بعدد من المرات = N ومقسومة على العدد N نفسه، كذلك فإن هذا الحد يساوي $\mu j2$ أو $Pj2$ والآن يمكننا إعادة كتابة معادلة تباين الفقرة على النحو الآتي:

$$P_j - 2P_j^2 + P_j^2 = \sigma_j^2 \qquad 2\text{-}5$$

وبجمع الحدود المتشابهة تصبح المعادلة:

$$P_j - P_j^2 = \sigma_j^2$$

وبإخراج العامل المشترك (Pj) تصبح المعادلة:

$$P_j(1 - P_j) = P_j^2 q_i = \sigma_j^2$$

وهكذا، فعندما تكون الفقرة ثنائية الدرجة فإنه يمكن حساب تباينها باستخدام المعادلة (5-3) وعند استخدام هذه المعادلة في حساب تباين الفقرة (1) في جدول (5-3) فإن:

$$0.25 = .5 \times .5 = س_1^2$$

وتعد هذه الصيغة أبسط في حساب تباين الفقرات ثنائية الدرجة من الصيغة المستخدمة في حساب التباين بشكل عام. وأما الانحراف المعياري فإنه يحسب بالصيغة:

$$\sqrt{P_j\ q_j} = \sigma_j$$

معامل الارتباط:

عندما يكون معد الاختبار مهتماً بالعلاقة بين الإجابات على فقرتين مثل (k,J) فإن هنالك

صيغة بسيطة لحساب معامل الارتباط والتي تستخدم P و q. وتسمى هذه الصيغة بمعامل ارتباط فاي، وهي على النحو الآتي:

$$P\Phi = \frac{P_j k - P_j\, P_k}{P_j q_j - P_k\, q_k} \qquad 6\text{–}5$$

حيث تمثل P_{jk} النسبة المشتركة للطلبة الذين أجابوا على الفقرتين j و k إجابة صحيحة. وهذه المعادلة (معادلة 5-6) مشتقة مباشرة من معادلة حساب معامل ارتباط بيرسون. ولتبين كيفية استخدام المعادلة أعلاه يبين جدول (5-4) نسبة الطلبة لكل نمط استجابة عل الفقرتين (k,j)، إذ تؤشر العلامة (+) إلى الاستجابة الصحيحة والإشارة (−) إلى الاستجابة الخاطئة على الفقرة.

جدول (5-4): توزيع النسب المشتركة لأنماط الاستجابة على الفقرتين (k,j).

	الفقرة (1) −	الفقرة (1) +	الفقرة (٢)
0.40	0.30	0.10	−
0.60	0.20	0.40	+
	0.50	0.50	

الفقرة (j) −	الفقرة (j) +	الفقرة k
b	a	−
d	c	+

شكل (5-1): الصيغة العامة للجدول رباعي الخلايا الممثل لنسب الاستجابات المشتركة للتوزيعات التكرارية على الفقرتين k , i.

وتمثل الخلية العليا إلى اليمين نسبة الطلبة الذين أجابوا على الفقرة (1) إجابة صحيحة وعلى الفقرة (2) إجابة خاطئة. وأما القيمة في الخلية العليا إلى اليسار فتمثل نسبة الطلبة الذين أخطأوا في الإجابة على كلا الفقرتين... وهكذا. وعند حساب معامل ارتباط فاي للفقرتين (2,1) باستخدام المعادلة (5-6) نصل إلى:

$$0.41 = \frac{0.6 \times 0.5 - 0.4}{\sqrt{0.4 \times 0.6 \times 0.5 \times 0.5}} = P(Ø)$$

ومن الممكن أيضاً حساب قيمة معامل فاي باستخدام تكرارات الاستجابات. ويبين الشكل (5-1) تكرارات أنماط الاستجابة لزوج من الفقرات، إذ يمثل كل حرف في الخلايا عدد الأفراد الذين أجابوا على الفقرتين وفق نمط الاستجابة المشار إليه في الخلية. فعلى سبيل المثال يبين العدد في الخلية (2) عدد الطلبة الذين أجابوا على كلا الفقرتين إجابة صحيحة. ويعد الجدول (5-5) رباعي الخلايا لبيانات جدول (5-3) وللفقرتين (2,1). وباستخدام هذه البيانات يحسب معامل فاي للفقرتين على النحو الآتي:

$$0.41 = \frac{bc - ad}{\sqrt{(a + b)\ (b+d)\ (C+d)\ (a+c)}} = P(Ø)$$

$$0.41 = \frac{2x1 - 4 \times 3}{\sqrt{(3 + 2)\ (2+4)\ (4+1)\ (3+1)}} = P(Ø)$$

وهكذا فإن درجة العلاقة بين فقرتين ثنائيتي الدرجة يتم حسابها باستخدام المعادلة (5-6) وبيانات صعوبة الفقرات، أو باستخدام المعادلة (5-7) وبيانات التوزيع التكراري للاستجابات بأنماطها الأربعة. وعند استخدام معادلة صيغة التكرارات على القارىء أن يتنبه إلى أن التوزيع التكراري مشابهٌ لما هو مبين في الجدول، وإن كانت الفئات مرتبة بشكل آخر فإن النتائج لا تكون صحيحة.

جدول (5-5): توزيع أنماط الاستجابة المشترك للفقرتين (1 , 2) من جدول (5-3).

	الفقرة (1) −	الفقرة (1) +
الفقرة (2) −	3	1
الفقرة (2) +	2	4

تباين المركب:

بما أن معظم الاختبارات تصحح بحساب درجات الفقرات وجمعها، ويتبع هذا أن توجد علاقة بين تباينات الفقرات المنفردة وتباين درجات الاختبار ككل. ويوجد في أسفل مصفوفة الفرد- الفقرة في جدول (5-3) «المتضمن استجابات (10) مفحوصين على اختبار مؤلف من (5) فقرات» خلاصةاحصائية متوسط وتباين كل فقرة، وعند حساب مجموع تباين الفقرات المنفردة نجد أنها تساوي (1.10)، ومن المدهش أن هذه القيمة أقل بكثير من تباين الاختبار الكلي ($o = 1.64$). فكيف نفسر هذه الفجرة؟ ومن الواضح أن تباين الاختبار الكلي لم يحسب من تباينات الفقرات المنفردة.

ولفهم سبب تجاوز تباين الاختبار الكلي لمجموع تباين فقراته منفردة، فمن الضروري فهم الصيغة العامة لتباين الدرجة الكلية (المركبة). لنفترض أن لدينا درجات الفقرتين ($x1$ و $x2$)، والمفحوصين نفسهم، من الممكن حساب مجموع درجات الفقرتين لكل مفحوص، ولنرمز لهما بالرمز C:

$$X_2 + X_1 = C \qquad (8\text{-}5)$$

ومتوسط الدّرجة المركبة للفقرتين هو :

$$\mu_2 + \mu_1 = \mu_2 \qquad (9\text{-}5)$$

ويمكن تحديد وسط المركب لهاتين الدرجتين من خلال الدرجات المنحرفة، وعلى النحو الآتي:

$$X_2 + X_1 = C \qquad (10\text{-}5)$$

$$\frac{\sum(X_1+X_2)^2}{N} = \frac{\sum C^2}{N} = \sigma_c \qquad (11\text{-}5)$$

ويمكن كتابة هذه الصيغة على شكل حدود من خلال تحليلها، وعلى النحو الآتي:

$$= \frac{\sum(X_1+X_2)^2}{N} = \frac{\sum X_1^2}{N} + \frac{\sum X_2^2}{N} + \frac{\sum 2X_1 X_2}{N} \qquad 12-5$$

ومن خلال التعريف، فإن الحدين الأول والثاني في المعادلة (5-12)، هما تباين الفقرتين، أي أن:

$$\frac{\sum X_1^2}{N} = \sigma_1^2 ، \quad \frac{\sum X_2^2}{N} = \sigma_2^2$$

ويمكن إعادة كتابة الحد الثالث في المعادلة (5-12) وعلى النحو الآتي:

$$= \frac{\sum 2X_1 X_2}{N} = \frac{2\sum X_1 X_2}{N} \quad \frac{\sigma_1 \ \sigma_2}{\sigma_1 \ \sigma_2}$$

وبإعادة ترتيب الحدود يمكننا كتابة:

$$= \frac{\sum 2X_1 X_2}{N} = \frac{2\sum X_1 X_2}{N} \quad \sigma_1 \ \sigma_2$$

وحيث أن الحد $\frac{\sum 2X_1 X_2}{N\sigma 1\sigma 2}$ عبارة عن معامل الارتباط $\rho 12$ يمكننا كتابة هذه الصيغة على النحو الآتي:

$$\frac{\sum 2X_1 X_2}{N} = 2P_{12}\sigma_1\sigma_2 \qquad 13-5$$

ويطلق على الحد $\rho_{12}\,\sigma_1\sigma_2$ اسم التباين المشترك بين X_1 و X_2. ولأي زوج من المتغيرات في العينة نفسها فإن التباين المشترك هو معامل ارتباط المتغيرين مضروباً بانحرافها المعياري. لنفترض الآن أننا أضفنا حداً ثالثاً للدرجة المركبة C، فإنها تصبح:

$$C = X_1 + X_2 + X_3$$

ووسط الدرجة المركبة المؤلفة من ثلاث فقرات يكون

$$\mu_y = \mu_1 + \mu_2 + \mu_3$$

وتكون صيغة التباين للدرجة المركبة المؤلفة من ثلاث فقرات على النحو الاتي:

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\rho_{12}\sigma_1\sigma_2 + 2\rho_{13}\sigma_1\sigma_3 + 2\rho_{23}\sigma_2\sigma_3 = \sigma^2 \quad (14-5)$$

وبالاستدلال نصل إلى استنتاج بأنه يكون هنالك حدوداً لتباين كل فقرة وضعف التباين المشترك بين كل فقرتين، وعلى هذا فإن الصيغة العامة لتباين الدرجة الكلية الناتجة عن مجموع درجات فقرات عدة تكون:

$$\sigma_1^2 + \sigma_2^2 + \ldots \sigma n^2 + 2\rho_{12}\sigma_1\sigma_2 + \ldots 2\rho_{nn-1}\sigma_n\sigma_{n-1} = \sigma^2_y \quad (14-5)$$

وستكون الصيغة أكثر دقة عندما تكتب على النحو الآتي:

$$\sum \sigma^2 + 2\sum_{i<j\ 1}\rho_{ij}\sigma_i\sigma_j = \sigma^2_y \quad (15-5)$$

وحيث أن مجموع الحد الثاني يمثل الأزواج المحتملة جميعها للتباين المشترك بين الفقرات المكونة للدرجة المركبة، لذا فإن التباين الكلي (تباين الدرجة الكلية) يحدد من خلال فقرات الاختبار والتباين المشترك بين أزواج الفقرات المحتملة جميعها.

إن أنسب طريقة لتصوّر عناصر التباين والتباين المشترك التي تساهم في تباين الدرجة الكلية من خلال مصفوفة التباين– التباين المشترك. وللاختبار المؤلف من فقرات عددها n فإن عناصر المصفوفة يساوي ($n \times n$)، وتقع تباين الفقرات على قطر المربع، بينما تقع عناصر التباين المشترك على أحد جوانب القطر، إذ أن القيمة في الصف (i) والعمود (j) تمثل التباين المشترك بين الفقرتين (j,i) (أنظر شكل (5-2)). ويمثل الرمز (σ_{ij}) صيغة مختصرة للتعبير الأكثر تعقيداً للتباين المشترك ($P_{ij}\,\sigma_i\,\sigma_j$). ويجب ملاحظة تماثل عناصر المصفوفة على جانبي قطر المصفوفة، بمعنى أن العنصر في الصف (i) والعمود (j) يماثل العنصر في الصف (j) والعمود (i) (أو إن $\sigma_{ji} = \sigma_{ij}$)

$$\begin{array}{ccccc}
\sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14}\ \ldots\ldots & \sigma_1 n \\
 & \sigma_2^2 & \sigma_{23} & \sigma_{24}\ \ldots\ldots & \sigma_2 n \\
 & & \sigma_3^2 & \sigma_{34}\ \ldots\ldots & \sigma_3 n \\
 & & & & \\
 & & & & \sigma_n^2
\end{array}$$

شكل (5-2): الصيغة العامة لمصفوفة التباين– التباين المشترك

لذلك فإننا نرى في المصفوفة فقط العناصر على أحد جانبي القطر وهكذا، فإن العبارة الجبرية $(2\rho_{ij}\ \sigma_i\ \sigma_j)$ تمثل $(\rho_{ij}\ \sigma_i\ \sigma_j + \rho_{ij}\ \sigma_i\ \sigma_j)$. ويمكن من الشكل استنتاج أن عدد عناصر تباين الدرجة المركبة هو n، في حين أن عدد عناصر التباين المشترك هو ((n(n-1)، فللاختبار المؤلف من خمس فقرات فإن المصفوفة تتألف من خمسة حدود للتباين وضعف هذا العدد لحدود التباين المشترك ($\frac{4 \times 5}{2} = 10$)

لنتفحص الآن جدول (5-6) الذي يتضمن مصفوفة التباين- التباين المشترك للاختبار المؤلف من الفقرات الخمس المبينة في جدول (5-3) السابق. إن فحص هذه القيم ومجموعها وكما هو مبين في الجدول أدناه يتبين لنا أن تباين الدرجات الكلي المساوي (1.60) هو مجموع تباينات الفقرات وتبايناتها المشتركة.

وأخيراً، ومن خلال فحص الصيغة العامة لمصفوفة التباين- التباين المشترك نجد أنه عند زيادة عدد فقرات الاختبار فإن عدد حدود التباين المشترك تزداد بشكل أكبر من عدد حدود تباين الفقرات. فعلى سبيل المثال، فعند إضافة (5) فقرات للاختبار المؤلف من (5) فقرات فإن عدد حدود التباين يزداد من (5 إلى 10) حدود، في حين أن عدد حدود التباين المشترك يزداد من (10 إلى 45)/ وفي بناء الاختبار فإن هذا له تطبيقات مهمة سيتم عرضها في الجزء الآتي:

جدول (5-6): مصفوفة للتباين- التباين المشترك للفقرات ثنائية الدرجة المبنية في جدول (5-3)

	1	2	3	4	5
الفقرة 1	.25	.10	.00	.00	.05-
الفقرة 2		.24	.06	.08-	.02
الفقرة 3			.24	.08	.08
الفقرة 4				.16	.06
الفقرة 5					.21

$$1.10 = \sum \sigma_1^2$$

$$0.54 = 2(0.27) = 2\sum_{i<J} \rho_{ij} \sigma_i \sigma_j$$

تضمينات عملية في بناء الاختبار

افترض أن باحثاً أعد صيغة تجريبية لمقياس قلق الاختبار والإجابة على فقراته إما أوافق أو لا أوافق، لذا فإن الاختبار يعد ثنائي الدرجة. ولهذا فإن مطور الاختبار عليه أن يفترض قلق الامتحان بناءً نفسياً وستكون الفروق الفردية مهمة عند تصميم الاختبار، وهذا يؤدي إلى تباين أساسي في درجات الاختبار. وبعد المحاولة الأولية لصيغة قصيرة من الاختبار (على سبيل المثال 20 فقرة) يجد الباحث أن التباين ليس كبيراً كما هو متوقع. ويتساءل الباحث عن كيفية تعديل الاختبار لزيادة التباين، ويأخذ بعين الاعتبار العوامل التي تؤثر في تباين الدرجةالكلية، فكيف يمكن هذا؟

1) هل يزداد تباين الاختبار إذا ازداد عدد فقراته؟ إن إضافة فقرات إلى الاختبار في أغلب الحالات يزيد التباين بقيمة تساوي تباين الفقرات ومجموع تبايناتها المشتركة مع الفقرات الأخرى الموجودة في الاختبار. إن نسبة الزيادة في تباين الاختبار الكلي ستكون ملحوظة بشكل أفضل عند إضافة الفقرات إلى الاختبار القصير أكثر مما لو أضيفت إلى الاختبار الطويل. ويمكن ملاحظة ذلك في الحالات الآتية: لنفترض إضافة (5) فقرات على اختبار مؤلف من (20) فقرة متوسط تباين فقراته (.20) ومتوسط تباينها المشترك (10) وتباين الاختبار الكلي (.42)، فعند إضافة (5) فقرات إليه لها متوسط التباين والتباين المشترك نفسه فإن تباين الاختبار الجديد يصبح:

$$.65 = (.10)\ (24)\ 25 + (.20)\ 25 = \sigma^2_x$$

والزيادة من (.44) إلى (.65) هي زيادة في تباين الاختبار الكلي وبما نسبته (55%). ومثال آخر لاختبار مؤلف من (70) فقرة بمتوسط تباين (0.2) لفقراته وتباين مشترك (0.10) وبتباين كلي للاختبار (.497)، وعند إضافة (5) فقرات أخرى للاختبار لها التباين والتباين المشترك نفسه يصبح تباين الاختبار الكلي على النحو الاتي:

$$.570 = (.10)\ (74)\ (75) + (.20)\ 75 = \sigma^2_x$$

والزيادة في التباين من (.497) إلى (.570) تمثل نسبة زيادة (15%) فقط من التباين الكلي. وعندما يؤخذ بعين الاعتبار تكاليف بناء وتطبيق الاختبار نصل إلى نقطة تكون عندها الزيادة الإضافية في تباين الاختبار الكلي غير متكافئة للنفقات الإضافية في المال والجهد

والوقت اللازمين. ويجب الإنتباه أيضاً إلى ارتباط الفقرات السلبي مع بعضها البعض لسبب ما فإن مثل هذه الفقرات ستقلل من تباين الاختبار الكلي بدلاً من زيادته.

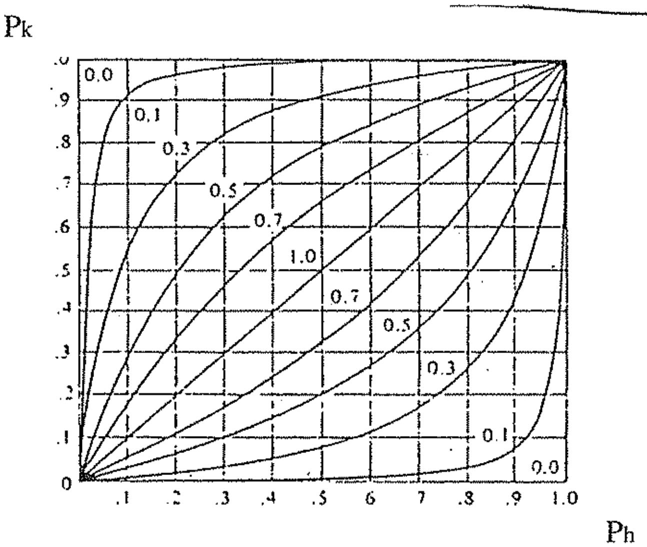
2) ما الخليط الأمثل لصعوبة الفقرات التي تزيد تباين الاختبار إلى أقصى قدر ممكن؟ يتضمن هذا السؤال جانبين. الأول: هل الأفضل إعداد فقرات متساوية في صعوبتها أم أن تكون لها صعوبات عالية ومتوسطة ومتدنية. وتذكر أن التباينات المشتركة للفقرة تكون أكبر عندما يكون ارتباط الفقرات أكبر، ومنطقياً تظهر الارتباطات الأكبر بين الفقرات عندما يجيب الأشخاص أنفسهم إجابة صحيحة على الفقرتين i و j، أو عندما تكون استجاباتهم خاطئة على كلا الفقرتين، وهنا تكون صعوبات الفقرات متساوية.

والوصف الأكثر دقة للعلاقة الرياضية بين معاملات ارتباط الفقرات وتساوي صعوباتها قدمه جوليكسن (Gulliksen, 1945) على شكل تمثيل بياني كالذي موضح في شكل (5-3). ويتبين من هذا الشكل أنه بزيادة الفجوة بين صعوبات الفقرتين (h ,k) فإن أقصى معامل ارتباط (بيرسون أو فاي) بين الفقرتين سينخفض بحدة، وكذلك يمكن تحقيق أقصى ارتباط (1.0) فقط عند تساوي صعوبات الفقرتين.

والثاني: ذكر جوليكسن أيضاً أن زيادة تباين الاختبار لا يكون فقط عند تساوي صعوبات الفقرات، وإنما يجب أن تكون متوسطة الصعوبة. ويبدو أن هذا سهل عندما نحسب تباين الفقرات من خلال قيم معاملات الصعوبة الآتية: (.20 , .40 , .50 , .70 , .90). ومن الواضح أن زيادة صعوبة الفقرة من أدنى قيمة (.20) إلى القيمة المتوسطة (.50) يزيد تباين الفقرة من (.16) إلى (.25)، وبعدها عند زيادة صعوبة الفقرة من (.50) إلى (.90) يبدأ تباين الفقرة بالإنخفاض من (.25) إلى (.09)، وهكذا فإن تباينات الفقرة وانحرافاتها المعيارية (والمساهمة في التباين المشترك) ستكون لها أكبر قيمة للفقرات متوسطة الصعوبة. ويجب التنبه إلى أن الاختبارات المؤلفة من فقرات من نوع الاختيار من متعدد أو الصح والخطأ الذي يظهر فيها أثر التخمين فمن الضروري أن تكون صعوباتها أعلى من (.50). وتحديد القيمة الأمثل لصعوبة هذه الفقرة ستناقش بالتفصيل في الفصل الرابع عشر.

3) هل يجب أن تكون الفقرات المضافة للاختبار مشابهة للفقرات الأصلية من حيث المحتوى، أم هل يجب أن تغطي محتوى جديد؟ إن تطوير الاختبار يتطلب إعداد سلسلة من الفقرات الأولية تغطي الأهداف أو مواصفات مجال محدد جيداً. ويمكن للقارئ تمييز أنه عند إضافة فقرات للاختبار فمن الأفضل إضافة فقرات سترتبط بدرجة أكبر مع الفقرات الموجودة في الاختبار، مما سيؤدي إلى زيادة التباينات المشتركة إلى

أقصى قدر ممكن. ومع ذلك فإن ترجمة هذا المبدأ في كتابة الفقرات قد يكون صعباً أحياناً. وفي محاولة تطوير فقرات جديدة للمحتوى نفسه فقد يتحفز معدّ الاختبار لتوسيع نطاق المحتوى الذي يغطيه الاختبار على أمل زيادة التباين الكلي أكثر من محاولته إعداد فقرات إضافية اعتماداً على المحتوى الأصلي فقط. والتطبيق الأول غير مرغوب به نوعاً ما، لأن هكذا فقرات إضافية قد تقلل من فائدة الاختبار لأنها تغير نطاق المحتوى الأصلي، ولأن معدّ الاختبار يخاطر بإنتاج فقرات غير مرتبطة بدرجة عالية مع الفقرات الأصلية (وهذا يساهم قليلاً في تباين الدرجة الكلية ثانياً). وهكذا ففي الحالات التي يقوم معدّ الاختبار بتطوير اختباره من خلال مجموعة مواصفات أو أهداف ينصح بإضافة فقرات على المحتوى نفسه فيما لو كانت الإضافة تهدف إلى زيادة تباين الدرجة الكلية.



شكل (3-5):

تجمعات معاملات ارتباط الفقرتين (Pk , Ph) التي لها أقصى ارتباط (1.0 , .7 , .5 , .1 , 0.00)

4) هل يُعد تباين الاختبار بحد ذاته مؤشراً على نوعية الاختبار؟ حتى الاختبارات التي تقيس الفروق الفردية، وتكون فيها هذه الفروق مهمة وأساسية فإن زيادة تباين الدرجة الكلية هو ليس الاعتبار الوحيد الذي يؤخذ بعين الاعتبار عند بناء الاختبار. وعند مقارنة اختبارين مصممين لخدمة الهدف نفسه فإن مستخدمي الاختبارات سيختاروا الاختبار الذي تكون درجاته لها صدق وثبات أعلى (وهذا ما ستركز عليه الوحدتان التاليتان)، وكما سنرى فإن هاتين الميزتين مهمتان جداً لدرجات الاختبار وتعتمدان على

تباين الدرجات. وهكذا فللاختبارات المعيارية فإن التباين ضروري لكنه ليس شرطاً كافياً لضمان فائدة الاختبار. بالمقابل فعندما يكون وصف المفحوصين في مجال ما هو هدف القياس (كما في الاختبارات المحكية) فإن بناء اختبار أو اختيار فقراته التي تزيد من تباين الدرجة الكلية ودرجة الفقرة يعد اعتباراً غير ملائم (أنظر على سبيل المثال (Popham, 1974)، (Popham & Husek, 1969)).

الخلاصة:

درجة الاختبار المركبة هي درجة الاختبار الكلية الناتجة عن جمع درجات اختبارين فرعيين أو أكثر. ودرجة الفقرة هي عدد النقاط التي تعطى لاستجابة المفحوص على فقرة معينة. وتحسب درجة الاختبار الكلية على أنها مجموع درجات الفقرات التي يمكن أن يشار لها على أنها الدرجة المركبة.

ويمكن تصنيف مخططات درجات الفقرة على أنها ثنائية (مقيدة بالقيم صفر وواحد) أو غير ثنائية. وللفقرات غير الثنائية تطبق الصيغ التقليدية للوسط والتباين والارتباط في وصف توزيع درجات الفقرات. وعندما تكون الفقرات ثنائية التفرع يمكن تطبيق صيغ خاصة في حساب الوسط والتباين وتستخدم هذه الصيغ مفهوم صعوبة الفقرة (P) الذي يعرف بأنه نسبة المفحوصين الذين أجابوا على الفقرة إجابة صحيحة. ويمكن حساب معامل الارتباط للمتغيرات (الفقرات) الثنائية باستخدام معامل فاي.

وعند اعتبار درجة الاختبار مركباً لدرجات الفقرات، فإن تباين المركب يساوي مجموع تباين الفقرات مضافاً إليه التباين المشترك للفقرات جميعها. ويساهم كل زوج من الفقرات (j, k) بتباين مشترك يساوي ($2\rho_{jk}\sigma_j\sigma_k$) ولاختبار عدد فقراته (n) يوجد تباينات عددها (n) وتباينات مشتركة عددها $\frac{n(n-1)}{2}$. لذا فعندما يكون الهدف تعظيم تباين الدرجة الكلية فإنه يمكن تقديم بعض المقترحات التي يمكن تطبيقها في بناء الاختبار:

1- زيادة عدد الفقرات يزيد تباين الدرجة الكلية.

2- يجب تساوي صعوبات الفقرات قدر الإمكان، وأن تكون مستوياتها في وسط المدى وذلك لتعظيم التباين المشترك للفقرات.

3- يكون الارتباط بين الفقرات متشابهة المحتوى نوعاً ما أعلى (وبالتالي تباين مشترك أعلى).
4- تباينات درجة الفقرة والدرجة الكلية لوحدها لا تعد مؤشرات لنوعية الاختبار.

التمارين:

1/ حصل خمسة تلاميذ على الدرجات المبينة أدناه في ثلاثة أسئلة مقالية:

آن	9	8	7
بيل	5	3	4
كارول	8	8	7
ديفيد	7	6	8
ايرين	6	5	7

أ- احسب μ و σ_i لكل فقرة.

ب- ما وسط درجات بيل.

ج- ما وسط درجات السؤال الثالث.

د- حدد تنوع الاستجابات التي مثلتها المجموعة على السؤال الأول.

هـ- ما الارتباط بين الفقرتين (2,1)؟ وعلى ماذا تدل هذه عن العلاقة بين الفقرتين؟

2/ كانت درجات (10) طلاب على اختبار رياضيات مؤلف من (5) فقرات على النحو الآتي:

	الفقرة				
الفرد	1	2	3	4	5
1	1	1	0	0	0
2	1	1	1	0	0
3	1	0	0	0	0
4	1	1	1	1	1
5	1	0	1	1	1
6	0	1	0	0	0
7	1	1	1	1	0
8	1	0	0	1	0
9	1	1	0	0	0
10	1	1	1	0	1

أ) أي الفقرات لها أعلى معامل صعوبة؟ وبماذا يخبرك عن الفقرة؟

ب) أي الفقرات لها أكبر تباين.

جـ- ما قيمة الانحراف المعياري للفقرة (3).

د) ما معامل الارتباط بين الفقرتين (5,4)؟

هـ- أعد جدول يبين بيانات الاستجابة على الفقرتين (2 و 3)، وباستخدام هذا الجدول احسب معامل الارتباط بين هاتين الفقرتين.

و) ما التباين المشترك بين الفقرتين (2 و 3)؟

3) تم الحصول على التباينات الآتية لبطارية مؤلفة من ثلاثة اختبارات فرعية:

$9 = \sigma_1^2$

$25 = \sigma^1_2$

$16 = \sigma_3^2$

إضافة إلى أن الارتباط بين الاختبارين الأول والثاني = .81 ، والأول والثالث = .64 ، والثاني والثالث = .90.

أ) أعد مصفوفة التباين – التباين المشترك.

ب) حدد تباين الاختبار المركب.

4/ أراد باحثان تطوير اختبار مؤلف من (10) فقرات لقياس دافعية الإنجاز. وكانت فقراته ثنائية التصحيح، وقد سمع الباحثون أنه يتم الحصول على أقصى تباين باستخدام فقرات متوسطة الصعوبة والتي ارتباطها عالي. لذا قرر هؤلاء تحديد فقرة واحدة لها تباين أساسي في الاستجابة، وسألوا المفحوصين عنها (10) مرات وذلك لأن هذه الاستجابات يجب أن يكون ارتباطها تاماً. فهل يعد هذا أسلوب منطقي في بناء الاختبار؟ وما النتائج المترتبة على هذا.

5/ برهن جبرياً على أن P_{ik} (معامل ارتباط بيرسون) بين المتغيرين (k , i) هو معامل فاي وذلك عندما تكون قيم (k , i) محددة فقط بالقيم صفر وواحد.

الوحدة الثانية

الثبات

الفصل السادس

6

الثبات ونموذج الدرجة الحقيقية التقليدي

الفصل السادس

الثبات ونموذج الدرجة الحقيقية التقليدي

عند بناء أي اختبار يفضل مستخدم الاختبار الحصول على النتائج نفسها عند تطبيق الاختبار على الأفراد أنفسهم في ظروف مشابهة. ويطلق على هذا الإتساق في درجات الاختبار إسم الثبات، وبمصطلحات تطبيقية هو المدى لانحراف درجات الفرد أو درجاته الزائية الذي تبقى معه متساوقة نسبياً عند إعادة الاختبار أو استخدام صيغ اختبارية بديلة. ولحد ما فإن الاختبارات النفسية جميعها غير ثابتة. فمثلاً إذا تقدم مجموعة لاختبار استعداد، ثم ارتاح هؤلاء مدة أسبوعين، فمن غير الممكن أن يحصل هؤلاء الأفراد على الدرجة نفسها في التطبيق أو أن يحتفظوا بالترتيب نفسه ضمن المجموعة. وبطريقة مشابهة، إذا تقدم طلبة الصف نفسه إلى صيغ بديلة لاختبار في الهندسة في اليوم نفسه، فمن المحتمل أن لا يحصل هؤلاء الأفراد على الترتيب نفسه في الاختبارين. إن مدى عدم ثبات مجموعة الملاحظات أو الاختبارات هي محط اهتمام كل من مطور الاختبار ومستخدميه.

ما الذي يجعل درجات الاختبار غير ثابتة؟ عندما يجيب المفحوص على مجموعة فقرات اختبارية فإن درجته تمثل فقط عينة محددة من السلوك- وهي استجابات على مجموعة فرعية من الفقرات المحتملة ضمن مجال معين، ثم الحصول عليها في واحد من المواقف الممكنة. لذا فإن الدرجات التي نحصل عليها تحت الظروف الاختبارية عرضة للخطأ وخاضعة لخطأ القياس. ويمكن تصنيف أخطاء القياس على أنها أخطاء عشوائية أو منتظمة. فالأخطاء المنتظمة هي الأخطاء التي تؤثر بانتظام على درجة الفرد بسبب خصائص معينة للفرد أو الاختبار التي لا نستطيع عمل شيء تجاهها للبناء المقيس. على سبيل المثال، في اختبارات القراءة للأطفال عندما يطلب الفاحص من الطفل وضع دائرة حول الحرف الذي يشير إلى صوت الحرف الأول في الكلمة، من المحتمل أن يسمع الطفل "bet" عندما يقرأ الفاحص كلمة "Pet"، عندها يشير الطفل إلى استجابة خاطئة. وعند إعادة الاختبار من الممكن أن يكرر الطفل الخطأ نفسه، وهنا تكون درجة الطفل منخفضة وباتساق عبر المواقف المختلفة. ويمكن توضيح الأخطاء المنتظمة للقياس باستجابة غير موافق من قبل مستجيب على مقياس اتجاهات عندما يقابل فقرة غامضة، ولأن مثل هذا الميل موجود عبر الاختبارات المتكررة للأداة نفسها، وتؤثر بنمط ثابت فإنها تعد أخطاء قياس منتظمة.

بالمقابل، تؤثر الأخطاء العشوائية للقياس على درجة المفحوص لأن حدوثها محض صدفة. فهي تؤثر على درجة المفحوص باتجاه موجب أو سالب، وتتضمن مصادرها: التخمين، والمشتتات في الموقف الاختباري، وأخطاء ادارية، وأخطاء تصحيح، وتقلبات في سلوكات المفحوصين، وتعد هذه كافية لتؤثر على الأداء الكلي في الاختبار (مثل تأثير الصداع على أداء المفحوص في الاختبار)، أو قد تكون مختصرة جداً ومحددة (مثل نسيان قراءة سؤال، أو نسيان كتابة مسألة رياضية، أو نسيان الإجابة). وأعد ستانلي (Stanely, 1971) قائمة تصنيفية مفصلة لمصادر التباين العشوائي الفردية. وعندما يعيد المفحوص الاختبار نفسه فإن الخطأ العشوائي الذي أثر على درجة المفحوص في التطبيق الأول قد لا يتكرر في التطبيق التالي، ومع ذلك فإن أخطاء عشوائية أخرى قد تظهر في المرة الثانية.

وكلاً من الأخطاء العشوائية والمنتظمة هي مصدر لاهتمامنا في تفسير الدرجات. فأخطاء القياس المنتظمة لا ينتج عنها عدم تجانس في القياس، ولكنها تبقى سبباً لعدم دقة الاختبار، وبالتالي تقلل من استخداماته العملية. وتقلل الأخطاء العشوائية كلاً من تجانس درجات الاختبار ومدى الإفادة منها. فمن غير المنطقي توقع فائدة من القياسات إن لم يكن لدينا ثقة بأنها متجانسة. لذا فمن مسؤوليات مطوّري الاختبارات البرهنة على ثبات درجات اختباراتهم وتتطلب مثل هذه البراهين دراسات تجريبية تعتمد في الأغلب على نموذج نظري يصف مدى تأثير الأخطاء العشوائية على الدرجات. وفي هذا الفصل سيتم وصف مثل هذا النموذج النظري الذي له تطبيقات واسعة في دراسة ثبات درجات الاختبار. وسنركز في الفصل التالي على الطرائق التطبيقية لدراسة الثبات الذي يعتمد هذا النموذج النظري.

نموذج الدرجة الحقيقية التقليدي:

قضية مهمة في هذا النموذج ترجع لمفهوم معامل الارتباط الذي جاء به عالم النفس الانجليزي شارلس سبيرمان. فمنذ عام 1904 وحتى عام 1913 كوّن سبيرمان دلائل منطقية ورياضية بأن درجات الاختبار عبارة عن قياسات معرضة لأخطاء الإنسان. وعلى هذا فإن الارتباط بين درجات الاختبار المعرضة للخطأ تكون أقل من الارتباط بين القيم الحقيقية للسمة المقيسة (Spearman, 1904). وبمحاولات متكررة لتفسير القياسات المعرضة للخطأ والقيم الحقيقية طرح سبيرمان (Spearman, 1907, 1913) أساس نموذج الدرجة الحقيقية التقليدي. وأعاد العديد من المؤلفين أمثال جيلفورد وجوليكسن وماغنسون ولورد ونوفيك صياغة هذا النموذج وتوضيحه، وكما هو موضح أدناه.

الأساس في نموذج سبيرمان أن أي درجة ملاحظة يمكن تصوّرها مؤلفة من عنصرين افتراضيين– الدرجة الحقيقية وعنصر الخطأ العشوائي. ويعبر عنها بالصيغة

$$X = T + E \qquad 1\text{-}6.....$$

حيث تمثل X الدرجة الملاحظة

و T الدرجة الحقيقية

و E عنصر الخطأ العشوائي.

على سبيل المثال، في اختبار مؤلف من (10) فقرات، يعرف جون في الحقيقة إجابات (7) فقرات ولكنه وضع إشارات خاطئة لفقرتين منها، لذا فإن درجة الملاحظة تصبح:

$$X = 7 - 2 = 5$$

وسارا تعرف إجابات (4) فقرات فقط، ولكنها بالتخمين أجابت على (3) فقرات إجابة صحيحة، لذا فإن درجتها الملاحظة تصبح:

$$X = 4 + 3 = 7$$

وأخيراً، يعرف رالف إجابة (8) فقرات، وخسر فقرة لعدم دقته في القراءة، ولكنه خمن وبشكل صحيح إجابة فقرة لم يكن يعرف إجابتها. هنا الخطأ الإيجابي حذف الخطأ السلبي، لذا فإن درجته الملاحظة:

$$X = 8 - 0 = 8$$

وتوضح هذه الأمثلة العددية تأثير الإضافة الموجبة والسالبة لأخطاء القياس، ولكنها غير صحيحة لنستدل منها على «الدرجة الحقيقية» للمفحوص كما حددها نموذج الدرجة الحقيقية التقليدي على أنها بصورة دقيقة هي عدد الفقرات التي يمكن أن يجيب عليها المفحوص.

درجات الاختبار كمتغيرات عشوائية

المتغير هو كمية يمكن أن يفترضها أي شخص لمجموعة من القيم. ويمكن تحديد المتغير العشوائي على أنه المتغير الذي تفترض قيمته وفقاً لمجموعة من الاحتمالات. فمثلاً، افترض أنك رميت حجر نرد (له ستة أوجه). يمكن اعتبار عدد النقاط التي تظهر على الوجه العلوي متغر عشوائي. ويمكن أن يأخذ هذا المتغير أي قيمة من (1 إلى 6) وفقاً للاحتمالات. ففي الرمية المفردة لحجر النرد تظهر فقط قيمة واحدة من القيم الستة، وهذه يطلق عليها «تحقيق المتغير العشوائي». ومع ذلك فإن هذا المتغير العشوائي له ستة قيم ممكنة. وعدد مرات تحقيق

المتغير العشوائي غير محدد، إذ أنه بإمكانك إعادة رمي حجر النرد مرة تلو الأخرى دون التأثير عليه. ويؤدي هذا إلى طريقة أخرى في التعبير المفاهيمي عن المتغير العشوائي. يمكن تصور المتغير العشوائي في مثالنا هذا على أنه توزيع افتراضي لنتاجات رمي حجر النرد، فالرمية الواحدة يمكن اعتبارها عينة عشوائية لنتاج واحد من توزيع النتاجات المحتملة. ومن المهم ملاحظة أن احتمال ظهور أي قيمة من القيم الستة غير معروفة، فلا يمكننا التأكيد على أنّ احتمالية ظهور أي قيمة هو $\frac{1}{6}$ حتى لو كان حجر النرد مصنوعاً بحيث يتسم بالعدل، فلا يوجد تأكيد على مثل هذه العدالة. وإذا حاولت تحديد هذه الاحتمالات تجريبياً برمي الحجر عدة مرات، وفي أحسن الظروف نحن نقدر الاحتمالية. وفي كل رمية جديدة للحجر تتغير التقديرات بدرجة قليلة.

وعندما يتقدم المفحوص لاختبار فإنه يمكن اعتبار درجته على الاختبار تحقيق للمتغير العشوائي. لماذا نطرح مفاهيمياً درجة الاختبار بهذه الطريقة؟ في البداية لاحظ أن الاختبار مؤلف من عدد محدد من الفقرات، فإن كان الاختبار مؤلف من (50) فقرة فإن درجة المفحوص يمكن أن تقع في أي مكان بين (1 إلى 50) وكما هو الحال في نتاج رمي حجر النرد إذ يتراوح المدى بين (1 إلى 6). ولا يمكننا قبل أن يتقدم المفحوص للاختبار معرفة القليل أو الكثير عن أمثلة قلة الإنتباه، والتخمين المحظوظ أو غير المحظوظ، وعدم قراءة الفقرة، وهكذا..

لذا يمكننا تصور درجة المفحوص في الاختبار على أنها قيمة محتملة لقيم عديدة لمجموعة احتمالات غير معروفة. وهذا التوزيع المحتمل لدرجات مفحوص واحد يمكن اعتبارها متغير عشوائي ، والدرجة التي يحصل عليها الفرد فعلاً عندما يأخذ الاختبار هي تحقيق لذلك المتغير العشوائي. ولتخيل كيف يمكن الحصول على هذا التوزيع الافتراضي لدرجات المفحوص تصوّر أن المفحوص تقدم للاختبار مرات عديدة، وبالإيحاء للمفحوص بنسيان الاختبار السابق، أعد تطبيق الاختبار مرات عديدة. بوضوح فإن الدرجات الملاحظة التي يحصل عليها الفرد في التطبيقات المتكررة للاختبار قد تنحرف بسبب أخطاء القياس المطروحة في بداية الفصل. ويمكن أن يزودنا التوزيع التكراري للدرجات التي حصل عليها الفرد بتقدير الاحتمالات التي تغطي درجات المفحوص في أي موقف اختباري. وفي هذه الحالة فإن الدرجة الملاحظة (مثل رمي حجر النرد) يمكن عدّها تحقيقاً للمتغير العشوائي.

ومن المهم تمييز أن درجة كل مفحوص في الموقف الاختباري تمثل متغير عشوائي مختلف، وهي أن احتمالية الحصول على درجة معينة تتحدد باستقلالية على توزيع مختلف لكل

مفحوص. ويوضح الجدول (6-1) هذه الحقيقة لمفحوصين تقدما لاختبار مؤلف من خمس فقرات (لاحظ أن هذا توضيح افتراضي إذ أنه لا يمكن ملاحظة هذه الاحتمالات مباشرة). ويمكننا اعتبار سوزان من خلال هذه التوزيعات المحتملة أكثر قدرة من إيلين، لأن هذه الاحتمالات تقترح أنها ستحصل على درجة أعلى من إيلين، وذلك عند النظر لقيمة واحدة من كل توزيع.

جدول (6-1): احتمالات درجات مفحوصين أثنين في عملية اختبارية واحدة.

درجة الاختبار (x)	احتمالية درجة سوزان	احتمالية درجة إيلين
0	0.00	.15
1	0.02	.20
2	.18	.40
3	.50	.23
4	.25	.02
5	.05	0.00

$$\sum_{\text{سوزان}} p(x) = (.00 \times 0) + (.02 \times 1) + (.18 \times 2) + (.50 \times 3) + (.25 \times 4) + (.05 \times 5)$$
$$= 3.13$$

$$\sum_{\text{إيلين}} p(x) = (.15 \times 0) + (.20 \times 1) + (.40 \times 2) + (.23 \times 3) + (.02 \times 4) + (.00 \times 5)$$
$$= 1.77$$

تعريف الدرجة الحقيقية:

وبعد أن رأينا كيف يمثل المتغير العشوائي التوزيع الاحتمالي، فمن الممكن التحدث عن متوسط توزيع مثل هذا. فالقيمة المتوقعة للمتغير العشوائي هي اسم آخر لمتوسط المتغير العشوائي. وعندما نفترض أن متغير عشوائي له عدد لا نهائي من القيم، فإن القيمة المتوقعة لـ x تتحدد كما يأتي:

$$\mu = \sum_{l=k}^{k} P_k X_k \quad \text{.....................} \qquad 2-6$$

حيث تمثل X_k القيمة k للمتغير العشوائي

و P_k القيمة المحتملة للمتغير العشوائي

وفي مثال حجر النرد (إن كان عادلاً تماماً) فإن القيمة المحتملة $\mu = \sum_{k=1}^{6} X_k P_k$

$$\mu = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + \ldots\ldots + 6\left(\frac{1}{6}\right) = 3.5$$

ومن الممكن افتراض عدد لا نهائي من القيم للمتغير العشوائي، وهنا لا تستخدم المعادلة (2-6) لتحديد القيمة المتوقعة، (إذ يحل محل الرموز الجبرية والعمليات المستخدمة طريقة رياضية تعرف بالتكامل)، وفي كلا الحالتين يرمز للقيمة المتوقعة بالرمز $\varepsilon(x)$، وعندما نعتبر الدرجة الملاحظة متغير عشوائي X_j، فإن الدرجة الحقيقية للمفحوص J تتحدد على أنها:

$$T_j = \varepsilon X_j = \mu x_j \quad \ldots\ldots\ldots \quad (3\text{-}6)$$

والقيم المتوقعة لدرجات المفحوصين في جدول (6-1). (وباستخدام معادلة 6-2) هي 3.13 لسوزان و 1.77 لإيلين. ويمكن اعتبار القيمة المتوقعة لكل مفحوص على أنها متوسط الدرجات التي حصل عليها في الاختبارات جميعها. وبتعبير غير دقيق يمكن اعتبار درجة المفحوص الحقيقية هي متوسط الدرجات الملاحظة التي حصل عليها من تطبيق الاختبار نفسه لعدد لا نهائي من المرات.

ومن المهم التمييز بين الدرجة الحقيقية على متغير نفسي والدرجة الحقيقية المطلقة لتغير فيزيائي أو بيولوجي. افترض على سبيل المثال أن الطبيب اشتبه بمرض كبد مزمن لمريض ما، فعلى هذا المتغير تكون للمريض درجة حقيقية مطلقة، فقد يكون مصاباً بالمرض أولاً. وحتى في حالة الدرجات الحقيقية المطلقة فإنه توجد احتمالية لظهور خطأ القياس. وهنا يستخدم الفحص المخبري للكشف عن هذا المرض، ومن الممكن أن يُعطي نتائج مختلفة عندما يُعاد على المريض نفسه. ومع ذلك فإن الدرجة الحقيقية المطلقة للمريض توجد وباستقلالية عن نتائج هذه الاختبارات. ولا يمكننا القول بأن حالة كبد المريض تتحدد بمتوسط قيم نتائج الاختبارات. علاوة على ذلك، وبغض النظر عن أنواع الاختبارات المختلفة التي أجريت فإنه يبقى للمفحوص درجة حقيقية مطلقة واحدة على هذا المتغير.. وبالمقابل تعتمد الدرجة الحقيقة للمريض في اختبار نفسي كلية على عملية القياس المستخدمة. فأي أخطاء منتظمة أو مظاهر تحيز لاختبار معين تعزى للدرجة النفسية الحقيقية للفرد في ذلك الاختبار. لهذا فإن كانت الدرجة الحقيقية على مقياس وكسلر للذكاء تنخفض باتساق بسبب خلل سمعي أو صعوبة لغة، وعند قياس الذكاء بمصفوفات رافن المتتابعة، فإن الفرد سيحصل على درجات حقيقية مختلفة على كل اختبار (لأن اختبار روكسلر يحوي عنصر لغوي). بالمقابل، فإن الدرجة الحقيقية النفسية

عبارة عن مفهوم احصائي يعتمد على القيمة المتوقعة التي يحصل عليها الفرد من عملية قياس معينة.

تعريف الخطأ:

بالنسبة لنموذج الدرجة الحقيقية التقليدي فإن خطأ القياس هو الفرق بين درجة المفحوص الملاحظة ودرجته الحقيقية. فالخطأ في درجة المفحوص j (Ej) تتحدد بأنها:

$$Tj - Xj = \varepsilon j \qquad .... \quad (4\text{-}6)$$

ويعد الخطأ Ej متغيراً عشوائياً، إذ أنه الفرق بين xj (متغير عشوائي) و Ti (مقدار ثابت للمفحوص j). ومتوسط توزيع الخطأ للمفحوص j هو القيمة المتوقعة:

$$(Tj - Xj) \quad E = \varepsilon Ej = \mu_E \qquad (5\text{-}6).............$$

ولتبسيط هذا التعبير أكثر، نستهدم قانونين أساسيين للعمليات المتعلقة بالقيمة المتوقعة. الأول: هو أن القيمة المتوقعة للفرق بين متغيرين هو الفرق بين قيمهما المتوقعة لذلك يمكن كتابة المعادلة (5-6) على النحو الآتي:

$$\varepsilon Tj - \varepsilon Xj = \varepsilon Ej \qquad (6\text{-}6).............$$

الثاني: القيمة المتوقعة لثابت هي الثابت نفسه، ولمفحوص J فإن:

$$Tj - \varepsilon Xj = \varepsilon Ej \qquad (7\text{-}6)...............$$

وبما أن $\varepsilon Xj = Tj$ (معادلة 3-6)، فإن :

$$0 = Tj - Tj = \varepsilon Ej \qquad (8\text{-}6)...............$$

والتفسير غير الدقيق هو أن معدل درجات الخطأ لمفحوص في اختبارات عديدة يساوي صفر.

خصائص الدرجات الحقيقية ودرجات الخطأ:

من خلال التعريفات السابقة، يمكن اشتقاق مبادىء أساسية لنموذج الدرجات الحقيقية التقليدي والتي تسمى افتراضات النموذج. وهذه المبادىء هي:

1/ متوسط درجات الخطأ لأي مجتمع من المفحوصين = صفر

$$(0 = \mu_E)$$

2/ الارتباط بين الدرجات الحقيقية ودرجات الخطأ للمفحوصين

$= 0 \quad (\rho_{TE} = 0)$

3/ عندما يتقدم المفحوصين لاختبارين منفصلين (أو موقفين اختباريين للصيغة نفسها) على افتراض اختيارها عشوائياً من توزيعين مستقلين للدرجات الملاحظة الممكنة فإن الارتباط بين درجات الخطأ للاختبارين = 0 ($\rho_{E1E2} = 0$).

4/ وهذه المبادىء الثلاثة تصف الخصائص الأساسية للدرجات الحقيقية ودرجات الخطأ مما يمكننا من تطبيق نموذج الدرجة الحقيقية التقليدي في دراسة ثبات درجات الاختبار. وفيما يأتي وصف غير رسمي للأسس المنطقية لهذه المبادىء والقاريء المهتم باشتقاق المعادلات الرياضية يمكنه الرجوع إلى لورد ونوفيك (Lord & Novick, 1968, P. 37-38).

متوسط الدرجات الحقيقية والخطأ:

افترض أن مجموعة مفحوصين تقدموا لاختبار، وكما تعلمنا فلكل مفحوص j درجة حقيقية Tj هي متوسط درجاته عبر اختبارات عديدة على الأداة نفسها أو على أدوات متوازية. لهذا، نظرياً فإن للمفحوص مجموعة درجات ملاحظة ممكنة على الاختبار، لهذا فإن:

$$\varepsilon X_j = T_j$$

ومتوسط درجات المفحوصين جميعهم في المجموعة هي:

$$\varepsilon_j T_j = \mu_T$$

حيث تمثل j – الرمز السفلي على ε– القيمة المتوقعة لأفراد المجموعة جميعهم. لهذا يمكننا تحديد متوسط الدرجة الحقيقية للمجموعة μ_T على أنها:

$$\varepsilon_j \varepsilon X_j = \mu_T \quad (9\text{-}6)$$

ورمز التوقع المكرر مكافيء لقولنا أن متوسط الدرجات الحقيقية لمجتمع المفحوصين مساوٍ لمتوسط الدرجات الملاحظة جميعها للذين تقدموا للاختبارات المتكررة جميعهم، أو أن:

$$\mu_T = \mu_x \quad (10\text{-}6)$$

ولايجاد متوسط درجات الخطأ للمجموعة μE نتبع الخطوات نفسها:

$$\varepsilon_j \varepsilon E_j = \mu_E \quad (11\text{-}6)$$

ومع علمنا بأن E_j ε = 0

فإن $\mu_E = \varepsilon_j (0)$

ومع العلم أن القيمة المتوقعة للثابت هي قيمة الثابت نفسها، فإن

$$\mu_E = 0 \quad ....(12\text{-}6)$$

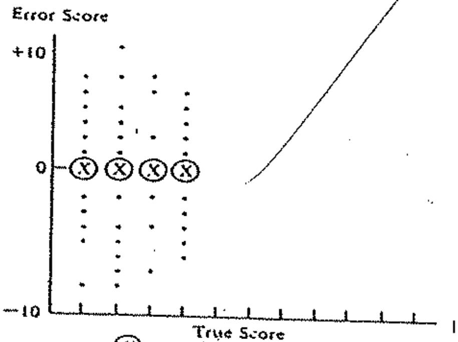
ومن المهم تذكر أن هذا المبدأ لا يعني أن متوسط درجات الخطأ للمفحوصين في اختبار معين يجب أن تكون صفراً. وتراكمياً، فإن E_j لمجموعة مفحوصين تتضمن مجتمع من الأخطاء بقيمة متوقعة (أو متوسط مجتمع) = صفر. وعند تطبيق اختبار واحد على مجموعة تتكون من 200 مفحوص فإن هذا يكافيء اختيار عينة لقيم E_j وباختيار قيمة واحدة عشوائياً من الدرجات الخطأ لكل مفحوص. ومتوسط درجات الخطأ لهذه العينة قد يساوي أو لا يساوي صفراً.

الارتباط بين الدرجات الحقيقية والدرجات الخطأ:

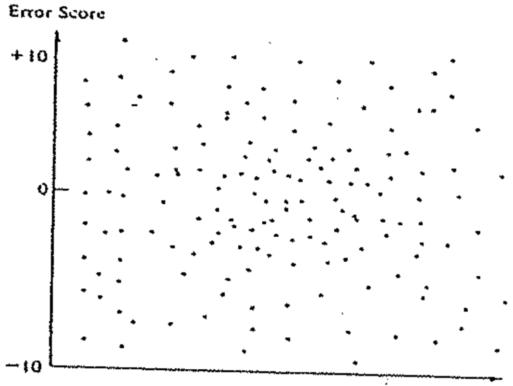
تخيل الموقف الذي نعرف فيه الدرجة الحقيقية ودرجات الخطأ جميعها ولكل مفحوص في المجتمع. سنستخدم هذه الدرجات في رسم مقطع بياني يمثل المحور العمودي فيه تدريج درجات الخطأ ويمثل المحور الأفقي فيه الدرجات الحقيقية. لاحظ أن لكل مفحوص عدد من درجات الخطأ ودرجة حقيقية واحدة، أي أنه يوجد عدد كبير من النقاط لكل قيمة من قيم الدرجات الحقيقية. افترض أننا وجدنا أدنى درجة حقيقية لأي مفحوص في المجموعة، وسنجد واحداً أو أكثر من المفحوصين عند هذه الدرجة. فلكل مفحوص في هذا المجتمع، وبالتالي لكل مفحوص لهذه الدرجة الحقيقية $\mu_E = 0$. وعلى هذا فإن متوسط درجة الخطأ لكل المفحوصين عند هذه الدرجة الحقيقية يساوي صفراً، وإذا حددنا درجات الخطأ جميعها لهذه الدرجة الحقيقية فإنها ستكون حول المتوسط صفر، بعدها افترض أننا وجدنا الدرجة الحقيقية الدنيا التي تليها وكررنا هذه العملية. ثانية ستكون درجات الخطأ حول المتوسط = صفر، وثانية ستكون الدرجات الخطأ حول المتوسط صفراً، وهكذا لكل درجة حقيقية، ويبين الشكل (6-1أ) رسم بياني لأدنى 5 درجات حقيقية T. ويظهر في الشكل (6-1 ب) قيم T الممكنة جميعها. ويمكننا من هذا التمثيل البياني رؤية أن الارتباط بين الدرجات الحقيقية والخطأ تساوي صفر وذلك لمجتمع الملاحظات الممكنة جميعها لكل المفحوصين. وهذا يعني أنه لا علاقة بين قدرة المفحوص وخطأ القياس والذي يؤثر على درجة المفحوص الملاحظة في أي موقف اختباري. وعلى هذا فإن التباين المشترك بين الدرجات الحقيقية ودرجات الخطأ = صفر، وعلى هذا فإن تباين الدرجات الملاحظة الكلي ببساطة هو مجموع تباين الدرجات الحقيقية وتباين الخطأ. ومعرفة هذه العلاقة تفيدنا لاحقاً في هذا الفصل.

الارتباط بين درجات الخطأ:

تخيل أن اختبارين طبقاً على مجتمع مفحوصين بأكمله، وأن ما نعرفه هو درجات الخطأ لكل مفحوص في كلا الاختبارين. افترض أن درجة الخطأ لمفحوص ما عالية وموجبة في الاختبار الأول، وإذا طلب منك أن تخمن درجة الخطأ لهذا المفحوص في الاختبار الثاني، فهل ستخمن أنها أيضاً موجبة وعالية؟ أم أنها سالبة (على افتراض أن أخطاء القياس توازن بعضها البعض، حتى على عدد محدد من الفقرات؟ الجواب: لا في كل حالة. تذكر أن كل درجة



(a) Plot for Selected True Score Values



شكل (6-1): أمثلة لرسم بياني للدرجات الحقيقية ودرجات الخطأ لمجموعة كبيرة من المفحوضين أجابوا على اختبار مؤلف من 10 فقرات

أن كل درجة خطأ للمفحوص في اختبار ما هي عينة عشوائية لواحدة من العديد من توزيعات درجات الخطأ ذات المتوسط صفر، ونتيجة للتعيين العشوائي لا توجد علاقة بين درجات الخطأ لحالتين اختيرتا من توزيعين مستقلين. ولأن هذا ينطبق على المفحوصين جميعهم، فإن التمثيل البياني الذي يبين نقاط أي تجمع ممكن لدرجات الخطأ من الاختبار الأول والثاني يكون مناسباً للشكل الموضح في (6-1ب).

لذلك نرى أن الارتباط بين الأخطاء لاختبارات معينة = صفر. فخطأ القياس العشوائي الذي يؤثر في درجة مفحوص في موقف لا يرتبط بخطأ قياس المفحوص نفسه في اختبار آخر.

دليل الثبات ومعامل الثبات:

بإعطاء تعريف الدرجة الحقيقية والخطأ، فإن الباحثين والمعلمين والمحللين النفسيين يعرفون فقط الدرجات الملاحظة للاختبارات التي يطبقونها، مع أن اهتمامهم الفعلي هو الدرجات الحقيقية. لهذا فإن سؤالاً مهماً يطرح حول كم هي العلامة وثيقة بين الدرجات الحقيقية والملاحظة؟ أحد مؤشرات هذه العلاقة هو الارتباط بين المتغيرين، ويسمى معامل الارتباط عن درجة العلاقة بين الدرجات الحقيقية والملاحظة «دليل الثبات» Reliability index.

تذكر أنه يمكن التعبير عن درجة المفحوص الملاحظة بالعلاقة:

$$X = T + E$$

وبالدرجات المعيارية:

$$x = t + e$$

وعند استخدام الدرجات المعيارية، يمكننا التعبير عن مؤشر الثبات على النحو الآتي:

$$\rho_{XT} = \frac{\sum t\,x^{(1)}}{N\,\sigma_X\,\sigma_T} \quad \cdots\cdots (6\text{-}13)$$

وبتعويض قيمة x، يصبح التعبير.

$$\rho_{XT} = \frac{\sigma_T^2}{\sigma_X\,\sigma_T}$$

أو

$$\rho_{XT} = \frac{\sum t^2}{N\,\sigma_X\,\sigma_T} = \frac{\sum t_e}{N\,\sigma_X\,\sigma_T} \quad \cdots\cdots (6\text{-}14)$$

ولأننا نفترض بأن الارتباط بين الدرجات الحقيقية والخطأ = صفر، فإن الشق الأخير من المعادلة (14-6) يحذف(2)، وحيث أن:

ان $\sigma_T^2 = \frac{\sum t^2}{N}$

فإنّ $\rho_{XT} = \frac{\sigma_T^2}{\sigma_T \ \sigma_X}$ .... (14-6)

والتي تبسط، وتصبح:

$$\rho_{XT} = \frac{\sigma_T}{\sigma_X} \quad \text{.......... (15-6)}$$

لهذافإنه يمكن التعبير عن دليل الثبات بأنه نسبة الإنحراف المعياري للدرجات الحقيقية إلى الإنحراف المعياري لدرجات الخطأ. لاحظ أن هذا الارتباط بين الدرجات الحقيقية والدرجات الملاحظة الممكنة جميعها لمواقف اختبارية متكررة. ويبدو أن هذا التطبيق له قيمة عملية قليلة لأنه لا يمكن الحصول على الدرجات الحقيقية مباشرة، كذلك لا يمكننا الحصول على الدرجات الملاحظة لكل مفحوص ومع ذلك من الممكن تخيل عملية اختبارية لمجموعة مفحوصين في موقفين باستخدام الاختبار نفسه أو صيغتين للاختبار نفسه. وعندما نجد اختبارين يحققا

---

1) بالتحديد نقول أنه من الخطأ استخدام رمز المجموعة ($\sum$) هنا لأن إشارة المجموع هذه تناسب فقط المجتمعات المحددة، ولا يوجد داعي هنا لتحديد المدى للمجتمع المحدد. في حين أن رمز التوقع (ε) أكثر عمومية، واستخدامه هنا يتطلب تفسير استخدام رمز التوقع المكرر، والذي قد يكون مربكاً للطلبة المستجدين. وقد ناقش لورد ونوفيك (1968) نموذج الدرجة الحقيقية التقليدي باستخدام أسلوب التوقع المكرر

2) بما أن $\frac{\sum t^2}{\sigma_E \ \sigma_X N} = \frac{\sum t_e}{\sigma_E \ \sigma_X} \left(\frac{\sigma_E \ \sigma_X}{\sigma_E \ \sigma_X} \right)$

وبإعادة تجميع هذا الحد يمكننا كتابته بالصورة:

$$= \frac{\sum t^2}{\sigma_E \ \sigma_X N} \left(\frac{\sigma_E \ \sigma_X}{\sigma_E \ \sigma_X} \right)$$

والان، تمثل القيمة الأولى في هذا التعبير ρ_{TE} والتي تساوي صفراً.

متطلبات التوازي من الممكن تأسيس رابطة رياضية ρXT: الارتباط بين الدرجات الحقيقية والملاحظة و $\rho xx'$: الارتباط بين الدرجات الملاحظة في اختبارين متوازيين. وفي نظرية الدرجة الحقيقية التقليدي يكون الاختباران متوازيين، إذا تحقق:

1/ لكل مفحوص الدرجة الحقيقية نفسها على صيغتي الاختبار.

2/ تساوي تباين الخطأ لكلا الصيغتين.

ونتيجة لذلك مثل هذه الاختبارات سيكون لها متوسطات متساوية وتباينات متساوية. ومن المنطق (ولكن ليس ضرورة احصائية) تطابق محتوياتها .ويمكن تحديد معامل الثبات على أنه الارتباط بين درجات الاختبارات المتوازية. وعندما نحدد معامل الثبات بهذه الطريقة فإنه يمكن وصف العلاقة بين دليل الثبات ومعامل الثبات رياضياً.

افترض أن الدرجات المعيارية لمفحوص على اختبارين متوازيين كانتا $x1$, $x2$، وبأسلوب هذا النموذج، يمكن التعبير عن الدرجات الملاحظة على النحو الآتي:

$$x_1 = t_1 + e_1$$

$$x_2 = t_2 + e_2$$

وصيغة الدرجات المنحرفة لمعامل الإرتباط بين الدرجات الملاحظة x_1 , x_2 هي:

$$\rho x_1 x_2 = \frac{\sum x_1 x_2}{N \sigma_{x1} \sigma_{x2}}$$

والتعويض بهذه المعادلة تصبح:

$$\rho\, x_1 x_2 = \frac{\sum (e_2+t_1)(e_2+t_2)}{N \sigma_{x1} \sigma_{x2}}$$

وبالحل الجبري لهذه الصيغة فإنها تصبح:

$$\rho\, x_1 x_2 = \frac{\sum (t_1+t_2)}{N \sigma_{x_1} \sigma_{x_2}} + \frac{\sum t_1+e_1}{N \sigma_{x_1} \sigma_{x_2}} + \frac{\sum t_2+e_1}{N \sigma_{x_1} \sigma_{x_2}} + \frac{\sum e_1+e_2}{N \sigma_{x_1} \sigma_{x_2}}$$

وضمن افتراضات نموذج الدرجة الحقيقية التقليدي فإن الحدود الثلاثة الأخيرة يمكن ملاحظة أنها تساوي صفراً، ولأنه يفترض تساوي قيم الدرجات الحقيقية عبر المواقف

الاختبارية جميعها، فإن $t_1 = t_2$ و $\sigma x1 = \sigma x_2$ (من تعريف الاختبـارات المتوازية، لذا فإن:

$$\rho x_1 x_2 = \frac{\sum t^2}{\sigma^2_{x_1} N}$$

أو

$$\rho x_1 x_2 = \frac{\sigma^2_T}{\sigma^2_x} \quad .... (16\text{-}6)$$

ومن هذه العلاقة نلاحظ أن معامل الثبات يمكن تحديده رياضياً على أنه نسبة تباين الدرجة الحقيقية إلى تباين الدرجة الملاحظة (أو أنه مربع دليل الثبات).

وفي تفسير معلومات الثبات، فمن المهم التمييز عبر:

1. $\rho x_1 x_2$ – نسبة تباين الدرجات الملاحظة التي تعزى لتباين درجات المفحوصين الحقيقية.
2. $(\rho x_1 x_2)^2$ – نسبة تباين الدرجات الملاحظة في أحد الاختبارات المتوازية التي يمكن التنبؤ بها من تباين الدرجات الملاحظة على الاختبار الموازي الثاني.
3. ρ_{XT} – معامل الارتباط بين الدرجات الحقيقية والدرجات الملاحظة. افترض أن دليل اختبار دوّن قيمة معامل ثبات ($\rho x_1 x_2 =$ .81) ويعد التفسير الآتي مناسباًلمعامل الثبات:

أولاً: يمكننا القول أن 81% من تباين الدرجات الملاحظة يعزى إلى تباين الدرجات الحقيقية للمفحوصين. لذا فإن $\rho^2_T = .81 \ \sigma^2_x$ وإذا كان الانحراف المعياري = 4، فإنه يمكننا التنبؤ بأن الإنحراف المعياري لتوزيع الدرجات غير الملاحظة يكون:

$$\sigma_T = \sqrt{16 \times 0.81} = 3.6$$

ثانياً: يمكننا القول أن $(\sqrt{.81})^2$ أو 65% من تباين الدرجات الملاحظة على الاختبار الثاني يمكن التنبؤ به من تباين الدرجات الملاحظة على الاختبار الأول.

وأخيراً: يمكننا القول أن الارتباط بين الدرجات الملاحظة والدرجات الحقيقية (يساوي .81 أو .90) لهؤلاء المفحوصين.

وعند هذه النقطة يجب ملاحظة أن معامل الارتباط لمجموعة المفحوصين هو مفهوم نظري بحت. فهو الكمية التي نحصل عليها إذا جزمنا بأن لدينا اختبارين متوازيين تماماً. وسنطرح

في الفصل السابع مشكلة استخدام درجات الاختبار الحقيقية في تقدير هذه الكمية النظرية. ولغاية هذه اللحظة سنشير ببساطة إلى أنه عندما نعوض بدرجات اختبار واقعية على أنها درجات اختبارات متوازية تماماً، فمن الممكن التصميم لجمع البيانات بطرق مختلفة. فقد نحاول التقريب في الحصول على الاختبارات المتوازية وذلك بتعويض مجموعة المفحوصين نفسها للاختبار نفسه في موقفين مختلفين. ومعامل الارتباط بين درجات المفحوصين نطلق عليه في هذه الحالة اسم معامل الاستقرار. أيضاً يمكننا محاولة تقريب القياسات المتوازية بتعريض المفحوصين لصيغتين مختلفتين للمحتوى نفسه وفي موقف اختباري واحد للمجموعة ذاتها، ويعرف معامل الارتباط بهذه الحالة على أنه معامل التكافؤ. كذلك يمكننا تعريض المفحوصين لصيغتين للاختبار نفسه في موقفين اختباريين منفصلين، وينتج عن ذلك معامل استقرار وتكافؤ. وهذه المعاملات جميعها وفي الأحوال المختلفة من الممكن أن يكون أقل من معامل الثبات النظري والذي نحصل عليه من قياسات متوازية حقيقية. وقد ميز كومبس (Coombs, 1950a) هذه الكمية النظرية على أنها معامل الدقة، وعرفها على أنها الارتباط بين درجات الاختبار عندما يستجيب المفحوصون لفقرات الاختبار نفسها مراراً وتكراراً دون حدوث أي تغيير للمفحوصين عبر الزمن، أو كما يفضل كرونباخ (Cronbach, 1951) وصفة بأن يكون الفاصل الزمني بين الاختبارين متلاشية (صغيرة للغاية).

وعندما يكون الهدف الأساسي لمطور الاختبار تقدير معامل الدقة لمجموعة درجات فعلية فإن معامل الإتساق الداخلي يحتسب أحياناً من درجات تطبيق واحد. وهنالك العديد من الطرق التطبيقية للحصول على معاملات الاتساق الداخلي المبينة في الفصل السابع. ولكن فهم الأساس لمعظم هذه الطرائق يعتمد على بعض المعرفة عن ثبات المركب لدرجات الاختبار.

ثبات المركب:

عرفنا المركب في الفصل الخامس على أنه درجة كلية تعتمد على درجات أثنين أو أكثر من الاختبارات الفرعية. دعنا نتخيل الآن أن مطور الاختبار ألف صيغتين متوازيتين رمز الأول A والآخر B، فإن كانت الصيغتان متوازيتين تماماً فإن معامل ارتباط أي منهما مع الآخر هو (ρ_{AB} ويستخدم كرمز شائع للثبات الرمز $P_{AA'}$ أو $P_{BB'}$. افترض أن مستخدم الاختبار يرغب بتطبيق كلا الصيغتين وحساب درجة كلية لكل مفحوص بالاعتماد على درجة المركب:

$$C = A + B$$

كيف يمكن تحديد ثبات الدرجات المركبة ($\rho_{CC'}$)؟ لاحظ أنه عند هذه النقطة يصبح تحايلاً اقتراح أن يبني مطور الاختبار صيغتين إضافيتين للاختبار ليزودنا بصيغة موازية لدرجات

المركب الأول. وحتى إذا فعل ذلك، فإن هذا السؤال حول ثبات المركب يبقى مطروحاً: ما هو ثبات المركب للاختبارات الفرعية الأربعة كلها؟ لذا فمن المهم إيجاد طريقة لتحديد ثبات المركب من خلال الخصائص الإحصائية للعناصر الداخلية. وفي هذا الجزء سنطرح طريقتين يمكن بواسطتهما التعبير عن ثبات المركب بوساطة الخصائص الإحصائية لمكوناته. تستخدم الطريقة الأولى ما يسمى نبوءة سبيرمان براون (Spearman Brown Prophecy)، وهذه تسمح بتقدير ثبات مركب اختبارات متوازية عندما يكون ثبات إحداها معروفاً. والطريقة الثانية تستخدم طريقة تعرف بكرونباخ ألفا، وهذه تسمح بتقدير ثبات المركب من معرفة تباين درجات المركب والتباين المشترك عبر عناصره أو مكوناته جميعها. ولتبسيط هذه المناقشة فإننا سنفترض البدء بمجموعة اختبارات مكوناتها متوازية. وبعدها سنطرح كيف تتأثر النتائج إذا لم تكن عناصر الاختبارات متوازية تماماً.

نبوءة سبيرمان براون:

لبدء عملية الاشتقاق، دعنا نعود إلى بعض العلاقات البسيطة التي تعلمناها (أو يمكن اشتقاقها منطقياً مما تعلمناه):

1/ يمكننا تحديد ثبات المركب على أنه $\rho_{CC'} = \dfrac{\sigma_{TC}^2}{\sigma_C^2}$ عادلة

2/ القياسات المتوازية جميعها لها متوسطات وانحرافات معيارية وتباينات متساوية، وعلاوة على ذلك أنه عندما يكون لدينا اختبارات متوازية عددها K، فإن الارتباط بين أي زوج منها يساوي الارتباط بين أي زوج آخر.

3/ إذا كان المركب مؤلّفاً من عناصر عددها k، فإن تباين هذا المركب يساوي مجموع تباينات العناصر وعددها k والتباينات المشتركة وعددها k(k-1).

وسنبدأ الآن اشتقاق صيغة ثبات المركب. دعنا أولاً نحدد المركب المؤلف من اختبارات متوازية عددها k.

$$C = A + B + \dots + k \qquad (17\text{-}6)$$

وتباين الدرجات الملاحظة للمركب هو:

$$\sigma_C^2 = \sigma_A^2 + \sigma_B^2 + \dots + \sigma_k^2 + \sum_{i=j} \rho_{ij}\, \sigma_j \sigma_L \qquad (18\text{-}6)$$

حيث تمثل $\sum_{i=j} \rho_{ij}\, \sigma_j \sigma_L$ مجموع التباينات المشتركة، وترمز i , j إلى أي زوج من الصيغ الاختبارية من A إلى k.

ولأن الاختبارات جميعها متوازية وقيم P_{ij} متساوية، لذا فإن:

$$\sigma_j + \sigma_i \ .... = \sigma_B + \sigma_A$$

وبالتالي يمكننا إعادة كتابة المعادلة (6-18) لتصبح:

$$[\sigma_{ij} \ (1-k) + 1\ ]\ \sigma_i^2 \ \ \sigma k = \sigma_2^2 \qquad (19\text{-}6).......$$

وأخيراً، لأن كل من i ، j قياسات متوازية، فإنه يمكن اعتبار P_{ij} معامل ثبات الاختبار i، ولذلك يمكن كتابة المعادلة (6-19) على النحو الآتي:

$$(\sigma_{ij} \ (1-k) + 1\ )\ \sigma_i^2 \ k = \sigma_2^2 \qquad (20\text{-}6).......$$

وسنستخدم هذا التعبير ليكون مقام ثبات المركب

والآن دعنا نعود إلى تباين الدرجات الحقيقية للمركب C والتي هي:

$$\sigma_{Ti} \ \ \sigma_{Ti} \ \ \rho_{Ti\,Tj} \ \sum_{j=i} + \sigma_{Tk}^2 + ...+ \sigma_{TB}^2 + \sigma_{TA}^2 = \sigma_{TC}^2 \qquad (21\text{-}6).........$$

ولأن الدرجات الحقيقية لكل مفحوص يجب أن تتساوى على القياسات المتوازية i , j، فإن $1 = \rho_{Tij}$ للاختبارات جميعها، ولأننا نتعامل مع الاختبارات المتوازية، فإن:

$$\sigma_{Tj} = \sigma_{Ti} = \sigma_{TB} = \sigma_{TA}$$

ولأن لدينا تباينات عددها k وتباينات مشتركة عددها k(1-k) في درجة المركب الحقيقية هذه، فإن:

$$\sigma_{Ti}^2\ (1-\ k\ ) + \sigma_{Ti}^2 \quad k = \sigma_{Tc}^2$$

وهذه يمكن تبسيطها أكثر لتصبح:

$$\sigma_{Ti}^2\text{-}k = \sigma_{Tc}^2 \qquad (22\text{-}6).......$$

وباستخدام صيغ σ_{Tc}^2 (من معادلة 6-22) و $\sigma^2{}_c$ (من معادلة 6-2)، يمكننا كتابة معادلة $\rho_{cc}\backslash$ على النحو الآتي:

$$\frac{\sigma^2{}_{Ti} \ \ k^2}{[\ P_{ij}\ ^{(1-K)} + 1]\sigma_i^2\ k}$$

ولأن $\sigma_{Ti}^2 \ / \ \sigma_i^2$ ، فإن هذا الكسر يمكن تبسيطه ليصبح:

$$\rho_{cc'} = \frac{P_{ij}\,k^2}{P_{ij}\,(1\text{-}k) + 1} \quad (23\ \text{-}6).......$$

وتعد المعادلة (6-23) الصيغة العامة لنبوءة سبيرمان براون، والتي تبين أن ثبات المركب يمكن التعبير عنه كدالة لثبات أحد مكوناته (على افتراض أن مكوناتها جميعها متوازية) وسيطرح في الفصل السابع الاستخدامات المهمة لهذه الصيغة في تقدير الثبات، وتطبيقاتها المطورة للاختبارات.

ثبات المركب بمعامل ألفا

الهدف الإجمالي لهذا الجزء هو إثبات أن ثبات المركب يمكن التعبير عنه بصورة دالة لتباين درجات مكوناته والتباينات المشتركة للاختبارات التي تؤلف بمجموعها المركب. وتعرف الصيغة التي سيتم اشتقاقها هنا بمعامل ألفا (Ctonbach, 1957)، والنقاط الآتية ستفيد القاريء في فهم عملية الاشتقاق:

1/ لأي زوج من الاختبارات المتوازية تماماً (j,i) يرمز للتباين المشترك بالرمز σ_{ij} أو

$$\sigma_{ij} = \rho_{ij}\,\sigma_i\,\sigma_j$$

2/ عندما يكون الاختبارين i , j متوازيين تماماً فإن تباين الدرجات الحقيقية لاختبار i يساوي التباين المشترك له مع درجات الاختبار الحقيقية j، أو:

$$\sigma_{Ti}^2 = \sigma_{TiTj}^2$$

3/ التباين المشترك الحقيقي لدرجات أي زوج من الاختبارات i , j يساوي التباين المشترك الملاحظ، أو

$$\sigma_{ij} = \sigma_{Tij}$$

وباستخدام هذه المعلومات، دعنا ثانية نجدد المركب C على أنه مجموع درجات الاختبارات الفرعية المتوازية:

$C = A + B + \dots + k$، كذلك فإن مركب الدرجات الحقيقية $T_C = T_A + T_B + \dots + T_k$. تذكر أن تباين هذه الدرجات:

$$\sigma_{TC}^2 = \sigma_{TA}^2 + \sigma_{TB}^2 + \dots + \sigma_{TK}^2 + \sum_{ij} \sigma_{TiTj}$$

حيث تمثل i , j أي زوج من الاختبارات الفرعية

و $\sum_{ij} \sigma_{TiTj}$ تمثل مجموع العناصر k (1-k).

ولأن القياسات k جميعها متوازية، فإن تبايناتها وتبايناتها المشتركة متساوية لأي زوج منها:

$$\sigma_{TC}^2 = k\,\sigma_{Ti}^2 + k\,k(1-k)\;\sigma_{Ti}\,\sigma_{Ti} \qquad (24\text{-}6)\ldots\ldots.$$

كذلك (من النقطة 2 في بداية هذا الجزء) نعرف أن

$\sigma_{Ti}^2 = \sigma_{TiTj}$ ، وبتعويض هذا في المعادلة (24-6):

$$\sigma_{TC}^2 = \sigma_{TiTj} + k\,k(1-k)\;\sigma_{TiTj}$$

وهذه يمكن تبسيطها إلى:

$$\sigma_{TC}^2 = k^2 \;\sigma_{ij\text{-}} \qquad (25\text{-}6)\ldots\ldots.$$

ومن النقطة 3 في بداية هذا الجزء $\sigma_{ij} = \sigma_{TiTj}$ وبتعويض هذا نحصل على:

$$\sigma_{TC}^2 = k^2 \;\sigma_{ij} \qquad (26\text{-}6)\ldots\ldots.$$

وإذا استخدمنا الصيغة الأخيرة على أنها تباين الدرجة الحقيقية للمركب، يمكننا كتابة الصيغة الآتية لثبات المركب:

$$\rho_{cc\backslash} = \frac{k^2 \;\sigma_{ij}}{\sigma_c^2} \qquad (27\text{-}6)\ldots\ldots.$$

وذلك عندما تكون الاختبارات جميعها قياسات متوازية.

وفي المواقف الاختبارية الواقعية لا نكون متأكدين من التوازي التام للاختبارات الفرعية التي تؤلف المركب. وفي هذه الحالة من الممكن استخدام مجموع التباينات المصاحبة وتباين المكونات لتقدير الحد الأدنى لثبات المركب. ويكون الحد الأدنى للثبات أقل من قيمة معامل الثبات، ولإشتقاقه علينا تكوين تباينات ثلاثة:

1/ عندما تكون الاختبارات الفرعية k غير متوازية تماماً، فإنه يوجد على الأقل اختبار فرعي واحد (وليكن g) له تباين درجة حقيقية أكبر أو تساوي التباين المشترك له مع أي اختبار فرعي آخر، أو أن

$$\sigma_{Tg}^2 \geq \sigma_{ij}$$

2/ لأي اختبارين غير متوازيين تماماً، فإن مجموع تباين الدرجات الحقيقي أكبر أو يساوي ضعف التباين المشترك k (1-k) مقسوماً على 1-k، أو أن

$$\sigma_{ij} \quad 2 \leq +\sigma_{Tj}^2 \, \sigma_{Ti}^2$$

3/ مجموع تباين الدرجات الحقيقية لاختبارات غير متوازية (عددها k) اكبر أو يساوي مجموع مجموع تبايناتها المشتركة k (1-k) مقسوماً على 1-k ، أو أن

$$\frac{\sigma_{ji^-} \sum_{j=i}}{1-k} \leq \sigma_{Ti}^2 \sum$$

والمبتاينة الأخيرة هي نتيجة التوسع المنطقي للنقطة 2 عبر اختبارات متوازية عددها k. وتفاصيل الخطوات الجبرية تجدها موضحة في (Lord Novick; 1968, P.89) وبإضافة مجموع التباينات المشتركة لطرفي المتباينة نحصل على:

$$\sigma_{ij} \sum_{j\neq i} + \frac{\sigma_{ij^-} \sum_{j=i}}{1-k} \leq \sigma_{ij} \sum_{j=i} + \sigma_{Ti}^2 \sum \qquad (28\text{ -}6).......$$

ويمكن تجميع مجموع التباينات المشتركة على الطرف الأيمن في حد واحد:

$$\frac{\sigma_{ij} \sum_{j\neq i} k}{1-k} = \frac{\sigma_{ij} \sum_{j\neq i} (1-k)}{(1-k)} + \frac{\sigma_{ij} \sum_{j\neq i}}{1-k}$$

إضافة إلى ذلك، فإن الطرف الأيسر للمتباينة هو تعبير عن σ^2_{Tc}، لذا فإن

$$\sigma_{ij} \sum_{j\neq i} \frac{k}{1-k} \leq \sigma^2_{Tc} \qquad (29\text{ -}6).......$$

وبما أن σ_{ij} هي مجموع التباينات المشتركة k (1-k) للاختبارات غير المتوازية تماماً، فإذا قسمنا طرفي المتباينة (6-29) على σ^2_2 ، نصل إلى:

$$\left(\frac{\sigma_{ij} \sum_{j\neq i}}{\sigma_c^2}\right) \frac{k}{1-k} \leq \frac{\sigma_{Tc}^2}{\sigma_c^2}$$

وهذه مساوية لـ :

$$\left(\frac{\sigma_i^2 \sum -1}{\sigma_c^2}\right) \frac{k}{1-k} \leq \rho_{cc\backslash} \qquad (30\text{ -}6).......$$

وتعرف الصيغة في الجانب الأيسر من المعادلة بمعامل ألفا.

للتلخيص، يتميز معامل الثبات النظري بأنه معامل للدقة (أي الارتباط الذي يمكن الحصول عليه بين اختبارين متوازيين تماماً إذا لم يحدث أي تغيير للمفحوصين في الفترة الواقعة بين الاختبارين) وعندما يكون الاختبار المركب مؤلفاً من اختبارات متوازية ولكن ليس تماماً، فإنه يمكننا تقدير الحد الأدنى لمعامل الدقة باستخدام معامل ألفا. ولحسابه يلزم معرفة عدد الاختبارات الفرعية، وتباين الدرجات، ومجموع التباينات المشتركة للاختبارات الفرعية: وتصبح أهمية هذه العلاقة أكثر وضوحاً إذا تذكرنا أن أي اختبار يمكن أن يكون مركباً، وأن كل فقرة فيه هي اختبار فرعي. وسنرى في الفصل السابع أن معامل ألفا يزودنا بطريقة ملائمة لتقدير الحد الأدنى لمعامل الدقة للاختبار وذلك باستخدام بيانات الإجابة على فقراته بتطبيق واحد.

الخطأ المعياري للقياس:

يعد الثبات مفهوماً يسمح لمستخدم الاختبار وصف نسبة التباين الحقيقي لدرجات الاختبار الملاحظة لمجموعة ما. وفي العديد من المواقف يكون مطور الاختبار مهتماً بدرجة أكبر بكيفية تأثير أخطاء القياس على تفسير درجات الأفراد، ومع ذلك فمن المستحيل تحديد كمية الخطأ بالضبط لدرجة ما. وتزودنا نظرية القياس التقليدية بطريقة لوصف التباين المتوقع للدرجات الملاحظة لكل درجة مفحوص حقيقية.

وتذكر أنه تم تعريف الدرجة الحقيقية على أنها المتوسط أو القيمة المتوقعة للدرجات الملاحظة عبر قياسات متكررة عديدة للاختبار نفسه. ويوضح الشكل (6-2) التوزيعات الممكنة للدرجات الملاحظة لمفحوصين حول درجاتهم الحقيقية ضمن التوزيع الأكبر لدرجات المجموعة الكلية على اختبار واحد. ويوجد انحراف معياري للمجموعة الكلية. ونظرياً، يوجد إنحراف معياري لكل توزيع درجات ملاحظة ممكن ولكل مفحوص حول درجته الحقيقية. وعندما تعدّل الإنحرافات المعيارية الخطأ للمجموعة الكلية فإن الناتج يسمى الخطأ المعياري للقياس ويرمز له بالرمز σ_E. ويمكن اشتقاق تعبير الخطأ المعياري للقياس باستخدام العلاقة:

$$\sigma_X^2 = \sigma_E^2 + \sigma_T^2$$

وبتقسيم طرفي المعادلة على σ_X^2 ينتج

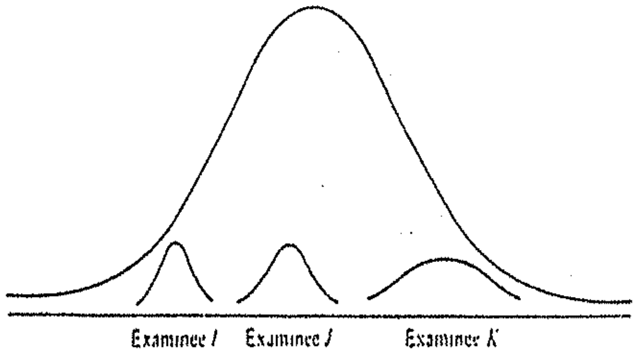
$$1 = \frac{\sigma_E^2 + \sigma_{T_i}^2}{\sigma_X^2 + \sigma_X^2}$$

لاحظ أن الحد الأول الأيمن هو تعريف الثبات ρxx، لذا فإن

$$1 = \frac{\sigma_E^2}{\sigma_X^2} + \rho_{XX\backslash}$$

ولحل المعادلة لاستخراج σ_E ، لاحظ أن:

$$\rho_{XX\backslash} \;\; -1 = \frac{\sigma_E^2}{\sigma_X^2}$$



شكل (6-2): توضيح افتراضي لتوزيعات الدرجات الملاحظة لمفحوصين مختلفين حول درجاتهم الحقيقية.

$$\sigma_E = \sigma_X \sqrt{1 - \rho_{XX\backslash}} \quad (6- 31)\ldots\ldots$$

فإذا كان الإنحراف المعياري للدرجات الملاحظة يساوي 10، ومعامل الثبات $\rho_{XX\backslash}=.91$ ، فإن قيمة الخطأ المعياري للقياس يمكن حسابها كما يأتي:

$$\sigma_E = 10 \sqrt{1 - 0.91} = 0.3$$

وعلى افتراض أن أخطاء القياس العشوائية تتوزع توزيعاً طبيعياً، فإننا نتوقع أن 68% تقريباً من الدرجات الملاحظة تقع في الفترة $T \pm 1\,\sigma_E$ ، و 95% من الدرجات تقع في الفترة $T \pm 1.96\,\sigma_E$ أو حول $T \pm 2\,\sigma_E$

وفي معظم المواقف الاختبارية يتقدم المفحوص لاختبار واحد ويكون له درجة ملاحظة واحدة. لذلك فإن معرفة الخطأ المعياري للقياس لا يمكننا من تكوين فترة فعلية حول درجة المفحوص الحقيقية، وذلك لأن القيمة الفعلية للدرجة الحقيقية غير معروفة. وبدلاً من ذلك فإننا

نستخدم القيمة المحسوبة للخطأ المعياري لتكوين فترة ثقة حول درجة المفحوص الملاحظة بصيغة $x \pm 1\,\sigma_E$ ، ويمكننا بنسبة ثقة 68% توقع الدرجة ضمن هذه الفترة. وتعويض الدرجة الملاحظة عوضاً عن الدرجة الحقيقية ضمن فترة الثقة يعد نوعاً من الخداع ولكن الحال ليس كذلك. افترض أن درجة جان الحقيقية على مقياس الإتجاه السياسي كانت (50)، وأن خطأ القياس المحسوب = 5 درجات. نظرياً إذا تم اختبار جان 100 مرة فإن 68 درجة من درجاته الملاحظة تقع 5 درجات حول درجته الحقيقية، أي بين 45 و 55، ولكن 32 درجة تقريباً: تكون خارج الفترة 45-55.

وإذا كوّنا فترات ثقة حول درجات جان الملاحظة (100 قيمة) فإن 68 قيمة منها تقع بين (45-55)، وكل من هذه الفترات تتضمن درجة جَان الحقيقية. وعملية اختبار جان مرة واحدة (كما هو في الأحوال العادية) فإنها تماثل أخذ إحدى درجاته منفردة وبشكل عشوائي من مجموعة الـ 100 درجة الملاحظة الممكنة. ففي الاختبار الواحد تكون الفرصة 68% للحصول على درجة ملاحظة تقع في الفترة (45-55). وعندما تكوّن فترة ثقة 5 درجات حول أي من هذه القيم، فإن هذه الفترة ستتضمن الدرجة الحقيقية 50. وإن كنا غير محظوظين فإننا سنحصل على إحدى الـ 32 درجة الملاحظة خارج الفترة (45-55). والفترة 5 درجات حول الدرجة الملاحظة سوف لا تتضمن قيمة الدرجة الحقيقية. ولهذا السبب فإنه من الضروري تذكر أن أي درجة ملاحظة مفردة قد تكون حساباً ضعيفاً لدرجة الفرد الحقيقية. ويعد الخطأ المعياري للقياس مهماً لتزويدنا بمدى بُعد الدرجة الحقيقية عن الدرجة الملاحظة لمعدل المفحوص ضمن المجتمع، ولكن لا يوجد إتفاق مطلق على أن درجة الفرد الحقيقية تقع ضمن فترة الثقة حول الدرجة الملاحظة. إضافة إلى ذلك يجب ملاحظة أن قيمة الخطأ المعياري تعكسσمتوسط الخطأ المعياري لمجموعة مفحوصين. ومن الممكن أن لا يكون منطقياً افتراض أخطاء معيارية متساوية للمفحوصين جميعهم. وستعرض إرشادات أكثر تحديداً لاستخدام الخطأ المعياري في تفسير الدرجات في الفصل السابع.

تعريفات بديلة للدرجة الحقيقية والخطأ:

حتى الآن، تعرف الدرجة الحقيقية على أنها متوسط عدد كبير من الدرجات الملاحظة التي حصل عليها المفحوص في مواقف اختبارية متكررة للاختبار نفسه أو اختبارات متوازية تماماً. ومع ذلك توجد نظريات بديلة للدرجات الحقيقية وأخطاء القياس التي تستخدم افتراضات أخرى. أحد هذه البدائل المعروفة تستخدم الموقف الاختباري الذي يُعد واحداً من احتمالات عدة لصيغ اختبارية تتألف من مجموعة كبيرة جداً من الفقرات. والقياس الذي نحصل عليه في هذه الحالات هو درجة المفحوص لصيغة اختبارية واحدة، ولكن يكون اهتمام

القياس هو تفسير أداء المفحوص على مجموعة الفقرات الكبيرة. وأعاد لورد (Lord, 1955, 1957, 1959b) تعريف الدرجة الحقيقية وأخطاء القياس في هذا السياق، والتي يمكن وصفها على النحو الآتي: افترض أن هنالك عددٌ كبيرٌ من الفقرات ثنائية الدرجة، يمكن باستخدام هذه الفقرات بناء اختبارين أو أكثر باختيار فقرات عشوائياً ضمن السياق، ومثل هذه الصيغ المتوازية ليس من الضرورة أن يكون لها متوسطات وتباينات متساوية، ولا أن يكون محتوى كل منهما مطابقاً للآخر، فدرجة المفحوص a الحقيقية هي عدد الفقرات التي يستطيع المفحوص الإجابة عنها بصورة صحيحة، ولكن من المناسب تعريف الدرجة الحقيقية على أنها نسبة الفقرات في المجموعة التي يستطيع المفحوص الإجابة عنها بصورة صحيحة (Pa). وإذا أردنا تحديد الدرجة الحقيقية للمفحوص a على اختبار طوله ثابت، حيث تمثل n عدد فقراته، فإنها تساوي:

$$n\,Pa = Ta$$

افترض أننا صممنا مجموعة فقرات عشوائياً لصيغة اختبار g. ومن الواضح أن g هي صيغة من عدة صيغ ممكنة يمكن تكوينها من مجموع الفقرات الكلي. وعلى هذا فإنه يوجد توزيع تكراري لقيم الاختبار الممكنة للمفحوص a ناتجة عن صيغ الاختبارات المختلفة والتي تتوزع عشوائياً حول Ta. ويعرف الخطأ المعياري للمفحوص a على أنه الإنحراف المعياري للتوزيع النظري لدرجات المفحوص a الملاحظة حول درجته الحقيقية.

لتطبيق هذه المفاهيم في موقف اختباري محدد، تخيل بنك فقرات آلي مؤلفاً من (1000) فقرة اختبارية تغطي محتوى مساق مدخل إلى علم النفس. وافترض أن المفحوص إليس يمكن أن يجيب عن (750) فقرة من هذا البنك، لذا فإن Pa = 0.75، وعندما يتقدم اليس للاختبار النهائي لصيغة اختبار مؤلف من (100) فقرة اختيرت عشوائياً من البنك، فإن درجته الحقيقية قد تكون:

$$75 = (.75)\,100 = nPa$$

وعلى الصيغة g يمكن أن يجيب اليس عن (80) فقرة إجابة صحيحة، وعلى الصيغة h قد يجيب على (73) فقرة إجابة صحيحة وهكذا. ونظرياً يمكن أن نحسب الأخطاء المعيارية لدرجات اليس بتكرار اختباره على الاختبارات الممكنة جميعها والمؤلفة من (100) فقرة، وحساب الانحراف المعياري لهذه الدرجات. وفي الواقع نحن لا نقوم بذلك وبدلاً من ذلك توجد طريقة عملية لحساب الخطأ المعياري للقياس تعتمد على حقيقة أن التوزيع التكراري لدرجات اليس للاختبارات الممكنة جميعها والمؤلفة من (100) فقرة لاختبارات متوازية عشوائية تقارب

توزيع ذي الحدين. وتوزيع ذي الحدين هو أحد أنواع التوزيعات التكرارية التي تعد مناسبة عندما نحسب احتمال رمي قطعة نقدية (50) مرة رأس و (1) مرة ذيل، أو (2) مرة رأس و (48) مرة ذيل، أو (3) مرة رأس و (47) مرة ذيل وهكذا. ونحن نعرف أن هنالك احتمالان فقط لكل رمية، ويوجد احتمال واحد كنتاج للرمية.والصيغة العامة للتوزيع ذي الحدين هي:

$$P(x) = \frac{n!}{x!\,(n-x)!} P^{x} Q^{n-x} \qquad (32\text{-}6)$$

حيث تمثل x (في مثال رمي القطعة النقدية) عدد مرات ظهور الرأس، و n تمثل عدد الرميات، و P تمثل احتمالية ظهور الرأس في إحدى الرميات، و Q تمثل احتمالية ظهور الذيل في رمية واحدة. هنالك ثلاثة خصائص للتوزيع ذي الحدين والتي تعد مهمة في الموقف الاختباري:

1/ يحسب متوسط توزيع ذ الحدين من الصيغة:

$$\mu = nP$$

2/ يحسب تباين توزيع ذي الحدين من الصيغة:

$$\sigma^2 = nPQ$$

3/ عند حساب تباين التوزيع من بيانات العينة، فإن صيغة تباين العينة يكون:

$$\hat{\sigma}^2 = \hat{n}\hat{P}Q \left(\frac{n}{n-1}\right) \qquad (33\text{-}6)$$

حيث أن $\hat{P}$ ، $\hat{Q}$ هي احتمالات مأخوذة من بيانات العينة، $\frac{n}{n-1}$ هي تصحيح لحساب تباين مجتمع غير متحيز وفي مثالنا، فإن توزيع درجات اليس المكنة تشبه توزيع ذي الحدين، لأن اختيار كل فقرة في صيغة اختبار تشبه رمية قطعة نقد. وعند اختيار كل فقرة فمن الممكن أن يستطيع اليس الإجابة عنها ومن الممكن لا. وعلاوة على ذلك فإن احتمال الإجابة الصحيحة عليها (.75)، واحتمالية خطأ الإجابة (.25). ودرجة اليس الكلية في الاختبار هو مجموع عدد الفقرات المختارة والتي استطاع الإجابة عنها إجابة صحيحة. وهذا مشابه مباشرة لعدد رميات قطعة النقد. وتكرار أي درجة كلية يمكن تحديدها من صيغة ذي الحدين في المعادلة 6-31.

وفي الواقع فإن اليس قد يتعرض لموقف اختباري واحد (صيغة واحدة)، وعلى هذا فإننا نعرف فقط الدرجة الملاحظة على هذا الاختبار (Xga) والنسبة الملاحظة الصحيحة Pga،

فكيف يمكن استخدام هذه المعلومات في تقدير الخطأ المعياري للقياس؟ إذا اعتبرنا نسبة الإجابات الصحيحة الملاحظة Pga تقدير لقيمة المجتمع Pa يمكننا استخدام المعادلة (6-32) في حساب تباين درجات اليس على صيغ الاختبار جميعها.

$$\hat{\sigma}_a^2 = n\ (\hat{P}ga\ .\ \hat{Q}ga)\ (\frac{n}{n-1}) \qquad \text{.......(6- 34)}$$

وبالاسترجاع لأن $\hat{P}ga = \frac{Xga}{n}$ و $\hat{Q}ga = 1 - \hat{P}ga$

نصل إلى أن:

$$\hat{\sigma}_a^2 = \sqrt{\frac{Xga\ (n - Xga)}{n-1}} \qquad \text{.......(6- 35)}$$

فالمعادلة (6-35) هي صيغة حسابية للخطأ المعياري للقياس للاختبارات المتوازية العشوائية. فالخطأ المعياري لدرجات اليس الذي حصل على الدرجة (80) على الصيغة (أ) المؤلفة من (100) فقرة قد تكون:

$$\hat{\sigma} = \sqrt{\frac{80\ (100 - 80)}{99}} = 4.02$$

ولمفحوص آخر حصل على الدرجة (50) في اختبار مؤلف من (100) فقرة اختيرت عشوائياً من البنك نفسه، فإن الخطأ المعياري المحسوب قد يكون 5.02. وتبين نتائج هذين المثالين تمييز مهم بين الخطأ المعياري ثنائي الحد للقياس وذلك المعتمد على نظرية القياس التقليدية. ففي نموذج الدرجة الحقيقية التقليدي نحصل عادةً على قيمة واحدة للخطأ المعياري، وعلى قيم مختلفة للخطأ المعياري ثنائي الحد المحسوب من درجات حقيقية مختلفة. وتكون الأخطاء المعيارية الثنائية أكبر للدرجات الحقيقية عند الوسط وأقل عند الدرجات المتطرفة (والخطأ المعياري ثنائي الحد يكون أكبر ما يمكن عند Pa= 0.50) وبسبب هذا التمييز فإن نماذج الخطأ المعياري قد تكون عملية أكثر من غيرها في بعض المواقف.

إن سياق نظرية القياس التقليدية هو قياس الفروق الفردية باستخدام الأداة نفسها في قياس المفحوصين جميعهم، ويكون الخطأ المعياري المعتمد على النظرية التقليدية مفيداً هنا، فعند استخدام اختبارات مختلفة لقياس الفروق الفردية (الاختبارات المحوسبة والتي تختلف باختلاف المفحوصين) أو للتفسير المطلق للدرجات (عوضاً عن التفسير المقارن) فإن الأنواع الأخرى للخطأ المعياري تكون مناسبة بدرجة أكبر. وللفقرات ثنائية التدريج يكون الخطأ المعياري ثنائي الحد مناسباً لكلا الموقفين المذكورين. إضافة إلى ذلك فإن النموذج ثنائي الحد

الذي طوره كيتس ولورد (Keats & Lord, 1962) يمكن استخدامه مع المشكلات التي لا يمكن حلها بنموذج الدرجة الحقيقية التقليدي. فمثلاً يمكن استخدام النموذج ذي الحدين في حساب درجة الاتساق بين تصنيف المفحوصين وفقاً لدرجاتهم الملاحظة والحقيقية. وهذا مهم جداً في الاختبارات المحكية. ويمكن استخدام الخطأ المعياري ثنائي الحد في اختبار ما إذا كان هنالك مقياسين يقيسان السمة نفسها. وهذا مهم جداً في فحص درجات الاختبار. فمثل هذه التطبيقات تحتاج لافتراضات أقوى من تلك المستمدة من النظرية التقليدية. ولهذا السبب فإن نموذج الدرجات الحقيقية التقليدي سيستمر استخدامه في مواقف عديدة تتطلب قياس الفروق الفردية لأنها تزودنا بحلول لمدى واسع من المشكلات القياسية والتي تحتاج لعدد قليل من الافتراضات.

الخلاصة:

يشير الثبات إلى اتساق أداء المفحوصين عبر التطبيقات المتكررة للاختبار نفسه أو الصيغ المتكافئة له. ومصدر رئيس لعدم اتساق الأداء على الاختبار هو أخطاء القياس العشوائية. ومن الضروري لمطوري الاختبارات ومستخدميها أن يحددوا مدى تأثير الأخطاء العشوائية على الأداء في الاختبار. ويزودنا نموذج الدرجة الحقيقية التقليدي بهيكل أو أساس نظري لتطوير دراسات ثبات تطبيقية.

ويعتمد نموذج الدرجة الحقيقية التقليدي على مفهوم أن درجة المفحوص في الاختبار عبارة عن متغير عشوائي، وتصوّر درجة المفحوص على أنها عينة عشوائية لواحدة من عدة احتمالات لدرجات الاختبار التي يمكن أن يحصل عليها الفرد عبر تطبيقات متكررة للاختبار نفسه (أو تحديداً صيغة متوازية لذلك الاختبار). ويؤشر لكل درجة اختبار ملاحظة على أنها مجموع الدرجة الحقيقية للمفحوص وعنصر الخطأ العشوائي. وتعرف الدرجة الحقيقية على أنها القيمة المتوقعة لدرجة المفحوص عبر مواقف اختبارية عديدة للاختبار نفسه. ويمكننا من خلال هذه التعريفات استنتاج ثلاث خصائص مهمة للدرجات الحقيقية والخطأ، هي:

1) متوسط الأخطاء لمجتمع المفحوصين يساوي صفراً.
2) الارتباط بين الدرجات الحقيقية والخطأ يساوي صفراً.

3) الارتباط بين الأخطاء في القياسات المختلفة يساوي صفراً.

وباستخدام نموذج الدرجة الحقيقية التقليدي يعرف معامل الثبات على أنه الارتباط بين القياسات المتوازية. ويمكن تبيان أن هذا العامل يساوي النسبة $\frac{\sigma_T^2}{\sigma_X^2}$: نسبة التباين الملاحظ الذي يعود إلى تباين الدرجة الحقيقي. ويسمى معامل الارتباط بين الدرجات الحقيقية والملاحظة لقياس واحد يسمى دليل الثبات (reliability Index) ويكافئ $\frac{\sigma_T}{\sigma_X}$

وباستخدام بيانات الاختبار يمكننا حساب معامل الثبات لإداة معينة بطرائق مختلفة وذلك بتطبيق الاختبار مرتين وينتج عن ذلك معامل الاستقرار. أو معامل التكافؤ أو معامل الاستقرار والتكافؤ. فالثبات المحسوب من تطبيق واحد نحصل عليه باستخدام صيغ ثبات المركب، وعندما نعتبر عناصر المركب متوازية تماماً يمكننا تعريف الثبات من خلال تباينات العناصر والتباينات المشتركة. وعندما يكون توازي العناصر غير تام فإن الحد الأدنى لمعامل ثبات المركب يمكن تحديده من خلال تباينات العناصر وتبايناتها المشتركة.

ويعرف الخطأ المعياري للقياس على أنه $\sigma_E = \sqrt{1-\rho_{xx}}$، ويمكن اعتباره متوسط الإنحراف المعياري لتوزيعات الخطأ لعدد كبير من المواقف الاختبارية المتكررة.

والخطأ المعياري للقياس مهم وذلك لإنتاج فترة ثقة حول درجات الاختبار الملاحظة التي لها احتمالية معروفة لتحتوي درجة المفحوص الملاحظة. والخطأ المعياري للقياس له قيمة لمجتمع المفحوصين في نظرية القياس التقليدية، وهو مهم في سياقات عديدة وذلك عندما نستهدف قياس الفروق الفردية. واقترح لورد (1955) أسلوباً بديلاً للخطأ المعياري للقياس لاستخدامه عندما تكون الصيغة الاختبارية ناتجة عن المعاينة العشوائية للفقرات من نطاق محدد بالضبط. ويعد الخطأ المعياري ثنائي الحد دالة لدرجة المفحوص الحقيقية وعدد فقرات الاختبار. ونتيجة لذلك فإن الخطأ المعياري هذا يختلف باختلاف مستويات قدرة المفحوصين. ومثل هذا الدليل يمكن تطبيقه في مواقف من مثل تقدم المفحوصين المختلفين إلى صيغ اختبارية مختلفة أخذت من ملف الفقرات نفسه.

التمارين:

1/ في كل من المواقف الآتية، بين ما إذا كان الحدث الموصوف يُعزى إلى خطأ قياس عشوائي أم منتظم في درجة المفحوص:

أ- قدر الملاحظ السلوك الإرشادي أثناء مقابلة المرشد والمسترشد الذي يميل لإعطاء الإناث درجات أعلى من الذكور على الفقرات المتعلقة بتقارير المسترشدين.

ب- أنزعج مفحوص يتقدم لاختبار رياضيات من الأصوات المزعجة ونتج عن ذلك خطأ حساب الجواب.

جـ- أصبح جين قلقاً أثناء الموقف الاختباري لدرجة أنها تركت فقرات عدة دون إجابة.

د- قفز مصحح لاختبار مقالي عن فقرة جعلت استجابة المفحوص غامضة.

هـ- مفحوص في صف علم نفس لم يكتب جملة من المحاضر، أدت إلى إجابته على فقرة في الاختبار إجابة خاطئة لأنها اعتمدت على المعلومات التي لم يدونها.

2/ تقدم أربعة مفحوصين لاختبار مؤلف من 3 فقرات، ويبين الجدول الآتي توزيع افتراضي لدرجات الاختبار للمفحوصين الثلاثة في تطبيقات متكررة للاختبار نفسه.

المفحوص	الدرجة			
	0	1	2	3
1	.5	.5	0	0
2	.25	.25	.25	.25
3	0	0	.5	.5
4	0	.5	.5	0

أ- ما الدرجة الحقيقية للمفحوص رقم 3.

ب- ما الدرجة الحقيقية للمفحوص رقم 2.

جـ- ما تباين الدرجة الحقيقية لهذا المفحوص.

د- ما تباين الدرجة الخطأ للمفحوص رقم 3.

هـ- ما تباين الدرجة الخطأ للمفحوص رقم 2.

و- ما معامل ثبات هذا الاختبار.

ز- فيما لو طبق هذا الاختبار في موقفين، أعطِ أمثلة لمجموعتين محتملتين للدرجات التي يمكن الحصول عليها لهؤلاء المفحوصين.

ح- في المواقف الاختبارية التطبيقية، هل يمكن تحديد ثبات درجات الاختبار بالطريقة المذكورة في و ؟ لماذا؟

3/ طبق باحث نفسي ثلاث صيغ متوازية لاختبار معياري على عينة البحث. وصححت الاستجابات باستخدام الراسم الضوئي. واكتشف الباحث فيما بعد أن هنالك خلل في وظيفة الراسم الضوئي نتج عن أخطاء في الدرجات على فترات عشوائية في اليوم الذي تم فيه تصحيح الأوراق، ولأن الباحث استخدم فقط متوسط درجة المفحوص عبر الصيغ الثلاث رأى أنه لا حاجة لإعادة تصحيح الاختبار، إذ أن أثر الأخطاء العشوائية تم التخلص منه إذ يمكن عدّ متوسطه صفر عبر الاختبارات الثلاثة. فهل هذا تفسير صحيح للإفتراض بأن متوسط أخطاء القياس يساوي صفراً؟

4/ تقدم جون لاختبار الاستعداد الأكاديمي الجمعي وحصل على درجة IQ المنحرفة قيمتها (135) نقطة. ثم تقدم بعدها لاختبار استعداد فردي وحصل على درجة IQ المنحرفة قيمتها 110 نقاط. ولاحظ استاذ جون أن الفجوة بين هذه الدرجات أكبر بكثير من الخطأ المعياري لأي من القياسين. ولم يدرك المشرف كيف حدث هذا، واندهش أيضاً حول أي القياسين يمكن عده قياساً أفضل لدرجة جون الحقيقية. كيف يمكنك تفسير هذا باستخدام نموذج الدرجة الحقيقية التقليدي.

5/ بين أن التباين المشترك للدرجات الملاحظة على اختبارين يكافيء التباين المشترك للدرجات الحقيقية للاختبارين، بصيغة أخرى اثبت أن:

$$\sigma_{T_1T_2} = \sigma_{X_1X_2}$$

(وتذكر ان $\sigma_{X_1X_2} = \frac{\sum X1X2}{N}$)

6/ أ- بين أنه للاختبارين المتوازيين i، j يكون

$$\sigma_{ij} = \sigma_{Ti}^2$$

ب- هل معرفة هذه العلاقة له تطبيقات عملية في تطور الاختبار.

7/ يرغب عالم نفس شخصية بتوليد المتغير Y وذلك بجمع درجة المتدرب الخام على مقياس دافعية التحصيل (X1) ودرجة الاستعداد الارشادي (X2). ووجد عالم النفس في دليل الاختبار معلومات عن الثبات والتباين لكل من X1 , X2 والارتباط بين X2 , X1. اشتق صيغة تعبر عن ثبات Y باستخدام هذه المعالم.

8/ يدون جدول (6-2) الارتباطات بين الأجزاء المختلفة للاختبارات الفرعية لبطارية التحصيل والمتوسطات والإنحرافات المعيارية، وثبات الصيغة البديلة للاختبارات الفرعية وذلك من بيانات طبقت على طلبة الصف الثاني في بداية النسبة الدراسية. استخدم هذا الجدول في الإجابة عن الأسئلة الآتية:

جدول (6-2): بيانات افتراضية في دليل بطارية اختبار تحصيلي لاستخدامها في تمرين 8.

الاختبار الفرعي	رقمه	1	2	3	4	5	6	7
الألفاظ	1	1.00	.66	.63	.68	.69	.65	.69
مهارات التعامل اللفظي	2		1.00	.87	.85	.61	.61	.55
الاستيعاب القرائي	3			1.00	.83	.57	.59	.53
اللغة	4				1.00	.62	.62	.55
مفاهيم رياضية	5					1.00	.72	.64
مفاهيم علمية	6						1.00	.61
دراسات اجتماعية	7							1.00
الانحراف المعياري		6.67	8.79	8.39	10.12	4.97	6.65	4.33
المتوسط		19.8	22.0	18.0	37.7	16.7	15.6	16.0
ثبات الصيغ المتكافئة		.87	.95	.95	.94	.81	.90	0.81

أ- ما معامل الارتباط بين درجات مهارة التعامل اللفظي والدراسات الاجتماعية.

ب- ما قيمة التباين المشترك بين هذين المتغيرين.

جـ-تتألف درجة القراءة الكلية من جمع درجات الاختبارات الفرعية: مهارات التفاعل اللفظي والاستيعاب القرائي. فما قيمة تباين الدرجة الكلية للقراءة.

د- ما معامل ثبات درجات القراءة الكلية.

هـ- كوّن عالم نفس معرفي متغير بإضافة درجات الألفاظ واللغة والدراسات الاجتماعية. فما قيمة معامل الارتباط بين المتغير الجديد ودرجة القراءة والكتابة.

و- لو حصل مفحوص على درجة 22 في الاختبار الفرعي للمفاهيم العلمية. ضمن أي فترة يمكن أن تكون درجة الثقة 68% لدرجة الحقيقة المتوقعة.

ز- من خلال افتراضات نظرية الدرجة الحقيقية التقليدي، كيف يظهر التمثيل بالأزواج المرتبة لدرجات الخطأ على اختبار الرياضيات والعلوم الفرعية؟

الفصل السابع

7

طرئق حساب الثبات

الفصل السابع

طرائق حساب الثبات

عرض الفصل السابق نموذج نظري يبين أثر الأخطاء العشوائية على درجات الاختبار. وفي نموذج الدرجة الحقيقية التقليدي يحدد معامل الثبات على أنه معامل الارتباط بين الاختبار المتوازية تماماً. كذلك يصور معامل الثبات على أنه المكافيء لنسبة التباين الملاحظ من درجات الاختبار التي تعزى إلى تباين درجات المفحوصين الحقيقية. وعملياً لا يمكن افتراض أن مطور الاختبار أعد اختبارين متوازيين تماماً، إضافة إلى أنه لا يمكن الحصول على درجات المفحوصين الحقيقية. لذا كيف يمكن الحصول على معامل الثبات لمجموعة من القياسات؟ الجواب هو أن معامل الثبات لمجموعة من الدرجات لا يمكن تحديده بالضبط ويمكن حسابه لعينة معينة من الأفراد أجابوا على عينة محددة من فقرات الاختبار. لاحظ أن مصطلح تقدير (estimate) لا تشير إلى النظر إلى البيانات ووضع تخمين حول الثبات ولكنه حساب عددي من عينة البيانات والتي هي تقدير للقيمة النظرية التي تم اختبارها. كذلك فإن استخدام الإشارة العليا " ^ " للرموز الإحصائية لكل من المتوسط، والتباين، والارتباط في هذا الفصل تشير إلى أن هذه الكميات محسوبة من العينة لا المجتمع بأكمله.

إن هدف هذا الفصل هو وصف الطرائق الشائعة المستخدمة في تقدير ثبات درجات الاختبار. ويمكن جمع البيانات المستخدمة في تقدير الثبات بطرائق عديدة. ويتوافر على الأقل وجهتا نظر لكيفية اختيار الطريقة الأنسب. إحداها هي الطريقة الأمثل التي تؤدي إلى حساب الارتباط الذي يمكن الحصول عليه إذا توافرت درجات اختبارين متماثلين تماماً. وهنا تصمم دراسة الثبات لتقليل أثر عدم توازي القياسات وذلك باستخدام مقاييس تكون أقرب إلى التوازي قدر الإمكان. والأخرى (والتي تشكل الأساس في هذا الفصل والفصل التالي) هي أن الطريقة الأكثر ملائمة التي يمليها الاستخدام الهادف لدرجات الاختبار. فيجب على مطور الاختبار أن يحدد مصادر خطأ القياس الأكثر ملائمة للتفسير المفيد للدرجات وتصميم دراسة ثبات تسمح بظهور مثل هذه الأخطاء وبالتالي تقييم أثرها.

طرائق تتطلب تطبيق اختبارين:

طريقة الصيغة البديلة:

افترض أن الخريجين الذي سيعملون في وظيفة معينة في المجال الصحي تقدموا جميعاً

إلى الاختبار العام (اختبار البورد) والذي طبق تحت شروط مضبوطة في موقف محدد في يوم محدد. ولخفض إمكانية الغش فقد أُعطي الطلبة الذين جلسوا على مقاعد متتالية صيغ اختبارية مختلفة تغطي المحتوى الاختباري نفسه. ومن حق كل مفحوص أن يتوقع عدم اختلاف درجته كثيراً بغض النظر عن الصيغة الاختبارية التي سيتقدم إليها. وفي هذه الحالة فإن خطأ القياس والذي يهتم به مستخدم الاختبار بصورة أولية يعود إلى اختلافات محتوى الصيغ الاختبارية المختلفة. كذلك فإن ظروف التطبيق وأخطاء التصحيح والتخمين والتقلبات الخاصة بأحوال المفحوصين قد تساهم أيضاً في عدم تطابق الدرجات. ومن أجل معالجة هذه العوامل على مطور الاختبار حساب معامل ثبات الاختبار باستخدام الصيغ البديلة.

وتتطلب هذه الطريقة بناء صيغتي اختبار متوازيتين وتطبيق كلا الصيغتين على مجموعة المفحوصين نفسهم. ويجب تطبيق كلا الصيغتين خلال فترة زمنية قصيرة جداً وتسمح بالوقت نفسه لفترة كافية بين التطبيقين لإلغاء الغموض عند المفحوصين. ومن المحبذ أيضاً موازنة ترتيب تطبيق الصيغ الاختبارية بحيث يتقدم نصف المفحوصين عشوائياً على الصيغة الأولى متبوعة بالصيغة الثانية، في حين أن النصف الآخر تطبق عليه الصيغة الثانية متبوعة بالصيغة الأولى. ويحسب معامل الارتباط بين مجموعتي الدرجات فيما بعد وباستخدام معادلة معامل ارتباط بيرسون، الذي يطلق عليه إسم معامل التكافؤ. وكلما كان معامل التكافؤ أعلى كلما كانت ثقة مستخدم الاختبار أكبر باستخدام الصيغ المختلفة بصورة متبادلة (كل مكان الآخر).

ولأي اختبار ذي صيغ عديدة يجب أن يتوافر فيها مؤشرات للتكافؤ فيما بينها. فالاختبارات التحصيلية واختبارات الاستعداد المدرسي تُبنى وبصيغ عديدة، إذ أن استخدامها في مجالات تحليلية أو علاجية وتربوية وبحثية يتطلب إعطاء فرصة للمفحوص لأن يتقدم للاختبار ثانية، وفي الوقت نفسه لا يرغب مستخدم الاختبار استعمال الفقرات نفسها عند تطبيق قوانين سهلة وسريعة للحد الأدنى المقبول لمعامل الثبات المحسوب. والعديد من أدلة الاختبارات التحصيلية المقننة تدون معاملات ثبات تتراوح بين (.80)و (.90) لهذا النوع من الثبات (التكافؤ)، بالإضافة إلى أنه يجب تدوين قيم المتوسطات والانحرافات المعيارية والأخطاء المعيارية للقياس لكل صيغة اختبارية. وقيمها يجب أن تكون متقاربة جداً فيما لو تم تفسير معامل التكافؤ على أنه قيمة تقديرية لمعامل الثبات.

طريقة الاختبار- إعادة الاختبار:

هنالك العديد من المواقف الاختبارية تكون فيها صيغة اختبارية واحدة كافية، ولكن اهتمام مستخدم الاختبار هنا هو كيفية استجابة المفحوصين على الصيغة نفسها في أوقات مختلفة.

وفي موقف كهذا تكون أخطاء القياس المهمة هي انحرافات درجات المفحوصين الملاحظة عن درجاتهم الحقيقية التي تعود للتغيرات غير الدائمة (المؤقتة) لحالة المفحوصين. وثانيةً فإن أخطاء التطبيق، والتصحيح، والتخمين، وعدم التأشير على الإجابة من قبل المفحوصين وكذلك الانحرافات المؤقتة في السلوك الأخرى قد تؤثر على الدرجات الملاحظة. ولحساب أثر مثل هذه الأخطاء على ثبات درجات الاختبار على مطبق الاختبار أن يختبر مجموعة مفحوصين، وينتظر ثم يعيد تطبيق الاختبار على المجموعة نفسها ثم يحسب معامل الارتباط الناتج عن تطبيق الاختبار – إعادة الاختبار والذي يطلق عليه أسم معامل الاستقرار.

ويرغب بمؤشر درجة الاستقرار في درجات المفحوصين في الاختبارات المستخدمة في تسكين المفحوصين في البرامج طويلة المدى. وهناك القليل من المعايير المتوافرة للحكم على أدنى قيمة مقبولة لمعامل ثبات الاختبار – الإعادة المحسوب. ومعاملات الثبات– الإعادة للاختبارات المنشورة متوافرة للاختبارات الاستعداد المطبقة فردياً. وعلى سبيل المثال متوافرة لمقياس الميول المهنية لسترونج (الطبعة الثانية)، فمعامل ثبات الصيغة المختصرة (80.) في حين أن معامل ثبات الصيغة الطويلة (60.). وكما نرى من النقاش أعلاه فإن العوامل المؤثرة في تقييم قيمة معامل الاستقرار المحسوب يجب أن تتضمن ذكر الفترة الزمنية بين التطبيقين وأعمار فئة المفحوصين والطبيعة النظرية للسمة المقيسة.

وسؤال مهم يطرح في تصميم دراسة ثبات الاختبار– الإعادة هو: ما الفترة الزمنية الفاصلة بين التطبيقين؟ لا توجد إجابة واحدة لهذا السؤال، فالفترة الزمنية يجب أن تكون كافية لدرجة تقلل من أثر الذاكرة أو التطبيق الأول على الأداء في المرة الثانية، وفي الوقت نفسه ليست طويلة تسمح بتغيرات النضج أو التاريخ ليظهر أثرها في درجات المفحوصين الحقيقية. ويجب أن يؤخذ بعين الاعتبار الهدف الذي ستستخدم فيه درجات الاختبار لتحديد فترة الانتظار بين التطبيقين. فعلى سبيل المثال ففي اختبار تقييم مستوى التطور النفسحركي للأطفال يوجد خطر قليل لتذكر الطفل استجاباته السابقة. وفي الوقت نفسه هنالك احتمال قوي لأن يؤدي النضج إلى تغيرات في الأداء فيما لو كان الوقت الفاصل بين التطبيقين طويل. وعلاوة على ذلك فمثل هذه الدرجات لا تستخدم في التنبؤ على المدى البعيد ولكنه يستخدم عموماً في التخطيط لبرامج فورية للتدخل الطبي أو النفسي والذي يُعاد تقويمه ثانية في فترة زمنية قصيرة. وعند أخذ هذه الأمور جميعها بعين الاعتبار فإن فترة انتظار من يوم لأسبوع تكون كافية. من جهة أخرى في استبانة الميول المهنية للبالغين يتذكر المفحوصون سريعاً بعض استجاباتهم على فقرات الاختبار وبعد فترات زمنية طويلة. ونظرياً، فإن الميول المهنية تكون مستقرة ولفترة زمنية طويلة إذ أن معظم البالغين يعملون في الوظيفة أو العمل نفسه ولفترة

زمنية طويلة. علاوة على أن درجات مثل هذه المقاييس تستخدم في إرشاد المفحوصين للإلتحاق ببرامج تدريبية تربوية أو مهنية التي تستمر لأسابيع عدة أو لسنوات لإكمالها، فإذا انحرفت ميول المفحوصين كثيراً خلال فترة التدريب كما هي مقاسة في هذه الأدوات فإن المعلومات الناتجة تكون ملائمة وتفيد مستخدم الاختبار.

وتعد الأمور التي تؤخذ بعين الاعتبار ليست مقنعة لجعل فترة الإنتظار بين التطبيق تتراوح بين (6) أشهر إلى سنتين. وبغض النظر عن فترة الانتظار بين التطبيقين فمن المهم تمييز أن معاملات ثبات الاختبار- الإعادة تختلف باختلاف الفترة الزمنية بين التطبيقين.

ولتفسير معامل الثبات على أنه تقدير للثبات يثير أسئلة مهمة: فعند الحصول على معامل ثبات منخفض فهل يعد هذا مؤشراً على أن الاختبار يقيس السمة قياس غير ثابت أم أن السمة نفسها غير مستقرة؟ فإذا إعتقد مستخدم الاختبار أن كمية السمة التي يمتلكها المفحوصون تتغير عبر الزمن فإن افتراض أساسي لنموذج الدرجة الحقيقية ينتهك ويكون معامل الارتباط المحسوب غير مناسب لتقدير ثبات درجات الاختبار. والقضية الثانية تتعلق بتغير سلوك المفحوصين بين التطبيقين الأول والثاني، ويعكس التطبيق الثاني في هذه الحالة أثر كل من الذاكرة والتعلم والملل والحساسية أو أية آثار أخرى للتطبيق الأول. وبوجود مثل هذه الأمور فمن المحتمل افتراض أن معامل الاختبار- الإعادة يمثل تقدير غير دقيق إلى حد ما لمعامل الثبات النظري. ومع ذلك فإن معلومات عن استقرار درجات الاختبار تكون خطرة لمستخدمي الاختبارات في مواقف اختبارية تطبيقية عديدة.

طريقة الاختبار- الإعادة بصيغ متبادلة:

يمكن حساب معاملات الثبات بالجمع بين طريقتي الاختبار- الإعادة والصيغ المتبادلة (المتكافئة). وهنا تطبق الصيغة الأولى للاختبار ثم فترة انتظار ثم تطبق الصيغة الأخرى. وإن كان بالإمكان فمن المرغوب به أن يكون ترتيب التطبيق لنصف المجموعة عكس الترتيب للنصف الآخر. ويعرف معامل الارتباط بين مجموعتي الدرجات على أنه معامل الاستقرار والتكافؤ. ويتأثر هذا المعامل بأخطاء القياس المرتبطة بمعاينة المحتوى في بناء صيغتي الاختبار وكذلك بتغيرات الأداء عبر الزمن وبالأخطاء المذكورة سابقاً جميعها. ويكون مثل هذا التقدير لمعامل الثبات في العادة أقل من كلا المعاملين التكافؤ أو الاستقرار المحسوب للاختبار نفسه وللمجموعة نفسها.

طرائق تتطلب تطبيق واحد للاختبار:

هنالك مواقف اختبارية عديدة تطبق فيها صيغة اختبارية واحدة على مجموعة المفحوصين. والمثال الأكثر شيوعاً لهذا النوع هو الاختبار الذي يُعده المعلم، فهو يعد صيغة اختبارية واحدة ويطبقها على المفوصين جميعهم. إضافة إلى أن المعلم لا يتوقع بالضرورة تطابق أداء الطلبة في الاختبار عبر الزمن إذ أن الطلبة سيتابعون تعلمهم أو سينسون المادة التعليمية وبمعدلات مختلفة. ويبقى الأكثر أهمية والملائمة الاهتمام بالمدى الذي يعكسه تباين الدرجات الملاحظ لتباين الدرجات الحقيقية في الاختبار، وفي وقت تطبيق الاختبار، وكما هو الحال في معظم المواقف الاختبارية فإن الفاحص لا يهتم بشكل أساسي بدرجات المفحوصين على الاختبار وفي العادة فإنه يستهدف التعميم من الأداء على الاختبار على مجال لمحتوى أكبر لفقرات محتملة يمكن أن تُطرح على المفحوصين ويختبر بها. واحدة من طرائق حساب اتساق أداء المفحوصين في الاختبار وإمكانية تعميمه على مجال فقرات أوسع هو تحديد كيفية وكم تطابق أداء المفحوصين عبر الفقرات أو المجموعات الفرعية من الفقرات في الصيغة الاختبارية الواحدة. والطرائق المصممة لحساب الثبات في هذه الظروف يطلق عليها اسم «طرائق الاتساق الداخلي». وطرائق الاتساق الداخلي المطروحة في هذا الفصل تعطي جميعها قيماً عبارة عن دوال لمعاملات الارتباط بين درجات جزأين منفصلين من الاختبار. وفي هذه الطريقة تؤدي طرائق عدة إلى قيم هي عبارة عن دوال الارتباط بين درجات نصفي الاختبار المصححين بشكل منفصل لكل نصف عن الآخر. ومن المهم التفكير في الارتباط بين مجموعات فرعية من الفقرات على أنها تزودنا ببعض المعلومات عن مدى إنعكاس بناءها للخصائص نفسها. فإن كان أداء المفحوصين متطابقاً عبر المجموعات الفرعية من الفقرات في الاختبار فإن الفاحص يكون لديه ثقة بإمكانية تعميم الأداء على فقرات أخرى محتملة في مجال المحتوى نفسه. وفي إجرائنا لدراسة الاتساق الداخلي فإننا نهتم بالأساس بالأخطاء الناجمة عن معاينة المحتوى، وكذلك الأخطاء الناجمة عن خلل في التطبيق أو التصحيح أو بسبب التخمين وتقلبات المفحوصين، وكذلك أثر البيئة التي يجري فيها الاختبار على أداء المفحوصين، وهذه العوامل جميعها قد تؤثر في معامل الإتساق الداخلي.

وعندما يكون أداء المفحوصين متسقاً عبر فقرات الاختبار يقال بأن الاختبار متجانس. وللحصول على مجموعة متجانسة من الفقرات فإنها يجب أن تقيس الأداء نفسه (أو أنها تمثل المحتوى نفسه. كذلك يجب أن تكون مكتوبة بشكل جيد وخالية من القصور التقني الذي يجعل المفحوصين يقدمون استجابات بناءاً على أسس معينة غير مرتبطة بالمحتوى. وعندما تؤخذ فقرات اختبار واحد من مجالات متنوعة (مثل الرياضيات والتاريخ والأدب) فإن أداء

المفحوصين لا يكون متسقاً عبر هذه الفقرات وبالتالي فإنها تؤدي إلى إنخفاض في معامل الإتساق الداخلي. وبصورة مشابهة فعندما تكون الفقرات من مجال واحد كالتاريخ على سبيل المثال ولكن بعض الفقرات تقيس مفاهيم أساسية وأخرى تعتمد على مفاهيم ثانوية مذكورة في حواشي الكتاب المقرر فقط فإن معامل الإتساق الداخلي يكون منخفضاً أيضاً.

أخيراً، إذا كانت الفقرات جميعها ممثلة بشكل غير متحيز لمجال المحتوى ولكن بعضها مكتوب بتعبير ضعيف لا يصف المطلوب من الفقرة فإنه يؤدي إلى تفسير خاطيء للسؤال أو أن تعتمد إجابة الفقرة على درجة الوعي بمضمون الفقرة لا على المعرفة، فهذه أيضاً تؤدي إلى خفض الإتساق الداخلي للاختبار. ومع ذلك فمن المناسب دائماً التحقق من معامل الإتساق الداخلي على أنه مؤشر لتجانس محتوى الاختبار ونوعية الفقرات.

وفي الجزأين التالين سنبين طريقتين واسعتي الإنتشار تستخدمان في حساب معامل الثبات من تطبيق واحد للاختبار. تعرف الطريقة الأولى بإسم طريقة التجزئة النصفية، والأخرى تستخدم تحليل التباين – التباين المشترك لاستجابات الفقرة. وكلا الطريقتين تؤدي إلى معامل تجانس داخلي لاستجابات المفحوصين على فقرات صيغة اختبارية واحدة.

طريقة التجزئة النصفية

يطبق مستخدم الاختبار ومطوره بهذه الطريقة صيغة اختبارية واحدة على المفحوصين جميعهم. وقبل تصحيح الاختبار يُجزيء مطور الاختبار الاختبار إلى مجموعتين أو اختبارين فرعيين يتألف كل جزء من نصف الاختبار الأصلي (من حيث الطول). لهذا فعند تطبيق اختبار مؤلف من (20) فقرة فإنه يُجزأ إلى نصفي اختبار كل منهما مؤلف من (10) فقرات. وهدف هذه التجزئة هو إنتاج نصفي اختبار متوازيين قدر الإمكان. وتوجد أربعة طرائق شائعة لتجزئة الاختبار إلى نصفين وهي:

1) اختيار الفقرات الفردية لتؤلف الاختبار الفرعي الأول والفقرات الزوجية لتؤلف الاختبار الفرعي الآخر.

2) ترتيب الفقرات على وفق صعوباتها المحسوبة من استجابات المفحوصين ثم اختيار الفقرات الفردية لتؤلف الاختبار الفرعي الأول والفقرات الزوجية لتؤلف الاختبار الفرعي الثاني.

3) اختيار فقرات كلا النصفين عشوائياً.

4) تجزئة الاختبار إلى جزأين بحيث يكون هنالك مزاوجة بين فقرات كلا الجزأين من حيث المحتوى.

وبعدها يصحح كلا النصفين مستقلاً عن الآخر ويحسب معامل الارتباط بين درجات

المجموعتين الفرعيتين من الفقرات، ويوضح الجدول (7-1) هذه الطريقة. ومفاهيمياً فإن معامل الارتباط هذا هو معامل تكافؤ نصفي الاختبار.

جدول (7-1): بيانات توضيحية لحساب معامل ثبات التجزئة النصفية

المفحوصين	الفقـــرة 1	2	3	4	5	6	المجموعة الفرعية أ الفردية	المجموعة الفرعية ب (الزوجية)	الدرجة الكلية
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1	0	1
3	1	0	1	1	1	0	3	1	4
4	1	1	1	1	1	1	3	3	6
5	1	1	1	1	1	1	3	3	6
6	0	0	1	0	0	0	1	0	1
7	0	0	1	1	1	0	2	1	3
8	0	0	0	1	0	0	0	1	1
9	1	0	1	1	1	0	3	1	4
10	0	1	0	1	0	1	0	3	3
المتوسط							1.6	1.3	2.9
الانحراف المعياري							1.28	1.19	2.02

$$.34 = \rho AB$$

$$.51 = .34 + 1/(.34)2 = \rho_{XX\backslash}$$

لاحظ أن معامل الارتباط الناتج عن استخدام هذه الطريقة يكون أقل من معامل الارتباط المحسوب للاختبار بأكمله (بشكل عام فإن الاختبار الأطول أكثر ثباتاً من الاختبار الأقصر لأن أخطاء القياس الناجمة عن معاينة المحتوى تنخفض). وللتغلب على هذه المشكلة يمكن لمطور الاختبار استخدام نبوءة سبيرمان- براون للحصول على القيمة المصححة لمعامل ثبات الاختبار الكلي. وعند تطبيق معامل الارتباط لنصفي الاختبار تكتب النبوءة على النحو الآتي:

$$\rho_{XX\backslash} = \frac{2P_{AB}}{1+P_{AB}} \quad .... (1-7)$$

حيث تشير $\rho_{XX}\backslash$ إلى معامل الثبات المعدل للاختبار الكلي و ρ_{AB} إلى معامل ارتباط نصفي الاختبار. لذلك فإن كان معامل الارتباط بين نصفي اختبار مؤلف كل منهما من (3) فقرات يساوي (.34)، فإن القيمة المصححة المحسوبة لمعامل الثبات للاختبار المؤلف من 6 فقرات تكون:

$$.0.51 = \frac{0.34 \times 2}{0.34 + 1} = \rho_{XX}\backslash$$

ويجب على مستخدمي معادلة سبيرمان- براون ملاحظة أن هذه الطريقة تعتمد بالأساس على افتراض التوازي لنصفي الاختبار. وكلما كان الإنحراف عن هذا الافتراض أكبر كلما كانت النتائج أقل دقة.

وطريقة بديلة لحساب الثبات من درجات نصفي الاختبار ودون استخدام نبوءة سبيرمان- براون اقترحها رولون (Rulon- 1939). وتتطلب طريقة رولون استخدام فرق الدرجات بين نصفي الاختبار:

$$B - A = D \qquad (2 - 7) \ldots\ldots\ldots\ldots$$

حيث تمثل A درجة المفحوص على نصف الاختبار الأول و B درجة المفحوص على نصف الاختبار الثاني. ويستخدم تباين فروق الدرجات σ^2_D على أنه حساب لتباين الخطأ σ^2_E في الصيغة التعريفية لمعامل الثبات، لذلك فإن:

$$\frac{\sigma^2_D}{\sigma^2_x} -1 = \hat{\rho}_{XX}\backslash \qquad (3 - 7) \ldots\ldots\ldots\ldots$$

وإذا حسب فرق الدرجة لكل مفحوص في جدول (7-1)، وتم حساب تباينها نحصل على حساب معامل الثبات وكما يأتي:

$$.0.45 = \frac{2.24}{4.08} -1 = \hat{\rho}_{XX}\backslash$$

ومع أن طريقة رولون هي الطريقة الأسهل استخداماً في الحسابات اليدوية فإنها تكافيء صيغة حساب الثبات بالتجزئة النصفية التي تعود إلى جوتمان (Guttman, 1945) أو فلانجان (Kelly- 1942).

وقد رأينا كيف نحصل على البيانات من تطبيق الاختبار لمرة واحدة وذلك بتجزأته إلى نصفين، ويمكن معالجة البيانات الناتجة عن التجزئة بإحدى الطريقتين:

1) حساب معامل الارتباط بين نصفي الاختبار وتصحيحه باستخدام نبوءة سبيرمان- براون.

2) يمكن استخدام درجات نصفي الاختبار لحساب ثبات الاختبار الكلي بطريقة رولون أو جوتمان.

ما مدى تشابه نتائج طريقتي سبيرمان- براون ورولون؟ عندما يتساوى تباين نصفي الاختبار فإن الطريقتين ينتج عنهما القيمة المحسوبة نفسها. إضافة إلى ذلك فقد بيّن كرونباخ (Cronbach, 1951) أن نسبة الإنحرافات المعيــارية لكلا النصفين عندما تتــراوح بين (.90) و(1.1) فإن كلا الطريقتين تعطي النتيجة نفسها، وعندما تزداد الإنحرافات المعيارية باضطراد فإن معامل سبيرمان- براون تكون قيمته أعلى وبانتظام من ذلك المحسوب بصيغة رولون.

وإحدى عوائق استخدام طريقة التجزئة النصفية سواء أكانت نبوءة سبيرمان- براون أم معادلة رولون فإنها لا تعطي قيمة واحدة لمعامل الثبات. وهناك طرائق عديدة ممكنة للتجزئة النصفية، وكلما لاحظ براونيل (Brownell, 1933) في مناقشته لهذه المشكلة أنه يمكن تجزئة الاختبار إلى نصفين بطرائق عددها $\frac{\frac{1}{2}\text{ك}\,!}{[(\frac{1}{2}\text{ك})\,!]^2}$ وذلك لاختبار عدد فقراته (ك). وينتج عن طرائق التجزئة المختلفة هذه معاملات ثبات مختلفة، وهذا يستحق الاهتمام.

طرائق تعتمد على التباين المشترك للفقرات:

إن الافتقار لطريقة واحدة في حساب التجانس الداخلي لدرجات الاختبار لعينة واحدة من المفحــوصين وفي تطبيق واحد، هذه المشكلة كــانت محط الاهتمام والأنظار في الأدب السيكومتري في الثلاثينات والأربعينات. وبعد ذلك وخلال فترة زمنية قصيرة اقترحت ثلاثة طرائق للتغلب على هذه المشكلة، ومع أن هذه الطرائق تبدو مختلفة في الصيغة إلا أنها تؤدي إلى نتائج متشابهة. والطرائق الثلاثة شائعة الإنتشار هي طريقة كيودر- ريتشاردسون وطريقة كروبناخ إلفا وطريقة تحليل التباين لهويت. وللذين يستخدمون أدلة الاختبارات (Test manuals) ومراجعات الاختبار (Test reviews) والأدب المتعلق بتطوير الاختبارات يجب أن يميزوا تكافؤ هذه الطرائق الثلاث ليتجنبوا التشويش، وسنستخدم في هذا الكتاب المصطلح معامل α ليشير إلى هذه الفئة من الطرائق.

معامل α : عرض كروبناخ عام (1951) تركيب شامل ومناقشة للطرائق المختلفة لحساب التجانس الداخلي وربطها جميعها بصيغة تعرف باسم معامل كروبناخ إلفا، وهي على النحو الآتي:

تستخدم كودر ريتشاردسون لاختبارات الاجابات الثنائية (٠، ١)
الفا كرونباخ مع اختبارات الاتجاهات والمقالية والاختبارات الثنائية
صورة تعتمد على تحليل التباين للأفراد والفقرات
لكل فقرة الطريقة لحساب معامل ألفا

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sigma^2_i}{\sigma^2_x}\right) \quad \text{.......... (4 - 7)}$$

حيث ترمز k إلى عدد فقرات الاختبار

و σ^2_i إلى تباين الفقرة i

و σ^2_x إلى تباين الاختبار الكلي

ويمكن استخدام صيغة ألفا (α) لحساب التجانس الداخلي للفقرات ثنائية التصحيح أو الدرجات ذات المدى الواسع من الدرجات (عديدة التدرج) مثل تلك في استبانات الإتجاهات أو الاختبارات المقالة. على سبيل المثال افترض أن المفحوصين اختبروا باختبار مؤلف من (4) أسئلة مقالية تراوحت درجاتها بين صفر إلى عشـــرة، وكـانت $\hat{\sigma}_1^2 = 9$ ، $\hat{\sigma}_2^2 = 4.8$ ، $\hat{\sigma}_3^2 = 10.2$ ، $\hat{\sigma}_4^2 = 16$ ، وتباين الاختبار الكلي = 100 ، فإن α تكون

$$\alpha = \frac{4}{3}\left(\frac{16 + 10.8 + 4.8 + 9 - 1}{100}\right) = 0.80$$

لتفسير هذه القيمة على أنها تقدير لمعامل الثبات يجب أن نلاحظ أن معادلة (7-4) هي ببساطة طريقة أخرى في التعبير عن معادلة ثبات المركب التي اشتقت في الفصل السادس. وفي سياق التطبيق الواحد للاختبار فإن كل فقرة تؤلف إحدى مجموعات ك الفرعية ودرجة الاختبار الكلية هي المركب. وكما بينّا في الفصل السادس فإن كانت لدينا الرغبة في افتراض أن الفقرات جميعها متكافئة تماماً فإن معامل ألفا يمثل حساب مباشر لـ $P_{XX'}$ للدرجة الكلية للاختبار. وفي معظم المواقف الاختبارية فإنه يتعذر الدفاع عن هذا الافتراض، لذلك نتقيد بالقول بأن $\rho_{XX'} \geq \alpha$ (نوقشت هذه المتباينة في الفصل السادس. وفي مثالنا كانت $\alpha = .80$.، يمكننا القول أن (80%) من تباين الدرجة الكلية على الأقل يعود إلى تباين الدرجة الحقيقية (أو أنه التباين المشترك الذي يؤثر بانتظام الأداء عبر الفقرات).

صيغتي كيودر- ريتشاردسون:

وكما لوحظ في بداية الفصل الحالي يمكن حساب معامل α ألفا بعدة طرائق تختلف عن الصيغة العامة المبينة في المعادلة (7-4). وإحدى الطرائق المعروفة الجيدة هي معادلة كيودر ... ريتشاردسون... (KR20)، التي يمكن استخدامها فقط للدرجات المصممة ثنائياً. وقد اشتقت هذه المعادلة من قبل كيودرو ريتشاردسون **(Kuder & Richardson, 1937)**. وكاستبصار لمشكلة التجزئة النصفية التي فشلت في إعطاء نتيجة واحدة لاختبار معين. والدراسة التي

قدماها كأرضية تتضمن معادلتين رياضيتين تعرف اليوم بـ (KR21 , KR20)، وأُخذت هذه الاسماء من الخطوات المرقمة في الاشتقاق المنشور في المجلة التي نشرت فيها المقالة. وصيغة KCR20 هي:

$$KR20 = \frac{k}{k-1}\left(1 - \frac{\sum \sigma q}{\hat{\sigma}^2_x}\right) \quad \text{.......... (5 - 7)}$$

حيث ترمز k إلى عدد فقرات الاختبار

و $\hat{\sigma}^2_x$ إلى تباين الاختبار الكلي

Pq تباين الفقرة في i

وهذه المعادلة مكافئة لمعادلة ألفا α عند تعويض $P_i\, q_i$ بـ $\sigma_i{}^2$ لاحظ أن المجموع يشير إلى أن تباين كل فقرة يجب حسابه أولاً ثم جمع تباينات الفقرات جميعها. ويبين جدول (7–2) حساب KR20 للاستجابات المبنية في جدول (7–1).

وبافتراض تساوي صعوبة الفقرات جميعها اشتق كيودر ريتشاردسون صيغة أبسط ولا تتطلب حساب تباين كل فقرة.

ويمكن كتابة صيغة KR21 على النحو الآتي:

$$KR21 = \frac{k}{k-1}\left[1 - \frac{\mu\,(k-\mu)}{k\,\sigma_x{}^2}\right] \quad \text{........... (6- 7)}$$

حيث ترمز μ إلى متوسط الدرجة الكلية.

$\sigma^2{}_x$ إلى تباين الدرجة الكلي.

و k إلى عدد فقرات الاختبار

ويبين جدول (7-2) أيضاً مثالاً لاستخدام صيغة KR21. وعندما تكون الفقرات متساوية في صعوباتها فإن كل من KR20 و KR21 ينتج عنها تقديرات متساوية للثبات. ومع ذلك فعندما تختلف صعوبات الفقرات فإن تقدير الثبات عند استخدام صيغة KR21 يكون أقل من القيمة المحسوبة باستخدام صيغة KR20. لهذا فمن غير المحبذ لناشر الاختبار أو الباحث أن يقدم تقدير KR21 للثبات لمجموعة درجات الاختبار. وللمعلمين ومستخدمي الاختبارات الذين يجرون الحسابات يدوياً أو باستخدام الحاسبة اليدوية فقد يجد معامل KR21 كافياً كحد أدنى لتقدير معامل الاتساق الداخلي لاختباراتهم.

طريقة هويت: طور هويت **(Hoyt, 1941)** وباستقلالية عن كيودر وريتشاردسون أسلوباً

لحساب معامل الثبات يؤدي إلى نتائج مماثلة لتلك الناتجة عن معامل ألفا. ويعتمد أسلوب هويت على تحليل التباين بمعالجة الأفراد والفقرات. على أنها مصادر للتباين.

جدول (7-2): حسابات توضيحية لمعامل الثبات KR وذلك لبيانات جدول (7-1).

الفقرة	P	q	Pq
1	.40	.60	.24
2	.30	.70	.21
3	.60	.40	.24
4	.70	.30	.21
5	.60	.40	.24
6	.30	.70	.21
			$\sum pq$ 1.35

$4.08 = \hat{\sigma}^2_x$ ، $2.02 = \hat{\sigma}_x$ ، $2.9 = \hat{\mu}$

$$0.76 = \left(\frac{(2.9-6)2.9-1}{(4.08)6}\right)\frac{6}{1-6} = KR21$$

$$0.80 = \left(\frac{1.35}{4.08} - 1\right)\frac{6}{1-6} = KR21$$

وباستخدام رموز تحليل التباين المعياري فإن معامل الثبات حدد على النحو الآتي:

$$\rho_{xx'} = \frac{MS\ persons - MS\ residual}{MS\ persons} \quad (7-7) ....$$

حيث تشير MS persons إلى مربع متوسطات الأفراد المأخوذة من جدول تحليل التباين.

و MS residual إلى مربع متوسطات البواقي المأخوذ من الجدول نفسه. ويبين جدول (3-7) خلاصة تحليل التباين للاستجابات المبينة في جدول (7-1)، ويوضح حساب معامل هويت. وقد ربط هويت هذه الصيغة بالتعريف النظري لمعامل الثبات. وذلك بملاحظة أن مجموع المربعات للأفراد (MS Persons) تمثل تباين الدرجات الملاحظة ومجموع المربعات للبواقي (MS residnal) تمثل تباين الخطأ في التعبير النظري للثبات.

$$\rho_{xx'} = \frac{\sigma_x^2 - \sigma_E^2}{\sigma_x^2}$$

جدول (7-3): جدول خلاصة تحليل التباين لهويت وذلك للاستجابات على الفقرات المبنية في جدول (7-1)

مصدر التباين	مجموع المربعات	df	مربعات المتوسط
الأفراد	6.817	9	.7574
الفقرات	1.483	5	.2967
البواقي	6.683	45	.1485

$$\rho_{xx'} = \frac{0.1485 - 0.7574}{0.7574} = 0.8039$$

ولأن طريقة هويت تتطلب معرفة طرائق إحصائية أكثر تعقيداً من الطرائق الأخرى لحساب معامل ألفا فإن القارئ سيندهش لذكرها هنا. والسبب الرئيسي لهذا هو أن تحليل التباين يعد طريقة إحصائية عامة أكثر إنتشاراً في الحقائب المبرمجة الاحصائية مما هو للبرامج المعدة خصيصاً لحساب معامل ألفا بأي طريقة أخرى. فهذه المعرفة توفر جهود البرمجة لمستخدمي الاختبارات الذين يرغبون باستخدام البرامج المحوسبة في حساب الثبات ومن خلال مجموعة واحدة من الاستجابات.

وكهدف للقارئ نلخص الأساليب الرئيسة لتقدير الثبات التي نوقشت في هذا الفصل، فإن جدول (7-4) يؤشر إلى مصادر الخطأ التي تؤثر على معامل الثبات المحسوب والطريقة المناسبة لجمع البيانات والتحليل المناسب للاستخدام في كل حالة.

جدول (7-4): ملخص لأساليب حساب معامل الثبات

المصدر الرئيس للخطأ	معامل الثبات	أسلوب جمع البيانات	المعالجة الإحصائية للبيانات P12^
1) تغير في المفحوصين عبر الزمن 2) معاينة المحتوى من صيغة لأخرى	1) معامل الاستقرار 2) معامل التكافؤ	1) اختبار- انتظار إعادة الاختبار 2) أعطِ الصيغة 1 ثم أعطِ الصيغة 2	1) احسب معامل ارتباط بيرسون $\hat{\rho}_{xx'}$ 2) احسب معامل ارتباط بيرسون $\hat{\rho}_{xx'}$

3) معاينة المحتوى أو قصور الفقرات	3) معامل اتساق داخلي	3) اعطِ صيغة واحدة في مرة واحدة	3) أ- جزيء الاختبار إلى نصفين، احسب الارتباط بين نصفي الاختبار ثم استخدم نبوءة سبيرمان براون. ب- جزيء الاختبار إلى نصفين واستخدم صيغة جوتمان أو رولون جـ- احسب تباينات الفقرات واحسب معامل ألفا

العلاقة بين تقديرات ألفا والتجزئة النصفية

من خلال التطورات الحادثة بين عامي (1935 و 1955) لطرائق تقدير ثبات الاختبار من تطبيق واحد، فمن الطبيعي ارتباك وحيرة مطوري الاختبارات حول الطريقة التي سيستخدمونها في حساب الإتساق الداخلي أو كيفية تفسير القيم المحسوبة (Coombs, 1950a). فمن الطبيعي أن يثير ظهور طريقة ألفا والطرائق المناظرة الأخرى جدلاً حول طبيعة ثبات الاختبار نفسه. فمن وجهة نظر المنظرين أمثال لوفنجر (Loevinger, 1947) الذين اعتقدوا أن تجانس الاختبار خاصية مهمة للاختبار على الرغم من وضوح المفهوم من وجهة نظر النظرية التقليدية لثبات الاختبار والذي يُعرف بأنه الارتباط بين صيغتين (أو نصفين) متوازيتين للاختبار. واستفسر كيلي (Kelley, 1942) عما إذا كان من المنطق الاهتمام بثبات إحدى صيغتي الاختبار فقط، ومن جهة أخرى أشار كورتون (Cureton, 1958) وغيره إلى أن الثبات يحدد بشكل مناسب على أنه نسبة تباين الدرجات الحقيقية إلى تباين الدرجات الملاحظة. ومن وجهة نظر هذا التعريف يبدو من المناسب اعتبار الثبات لصيغة اختبارية منفردة. وتاريخياً فإن وجهة النظر الأخيرة تم تأييدها من قبل كرونباخ (Cronbach, 1951) الذي وضح العلاقة بين طريقتي ألفا والتجزئة النصفية. ومناقشته تضمنت النقاط الآتية في التفسير المناسب لمعامل ألفا:

1) يمكن استخدام معامل ألفا على أنه معامل إتساق داخلي. أنه خاصية للاختبار تدعمها حقيقة الارتباطات الداخلية الموجبة بين الفقرات التي تؤلف الاختبار (Kuder & Richardson, 1937) وفي تفسير معامل ألفا يجب أن يتذكر مستخدم الاختبار أن هذا التقدير لا يؤشر إلى استقرار درجات الاختبار عبر الزمن أو التكافؤ مع درجات صيغة أخرى للاختبار.

2) يمكن اعتبار معامل ألفا الحد الأدنى لمعامل الثبات النظري ويُعرف باسم معامل الدقة، لذا فإنه ليس حساباً مباشراً للثبات ولكنه حساب للحد الأدنى لهذا المعامل، فإذا حصلنا على القيمة (75.) لمعامل ألفا فمن المستحيل معرفة ما إذا كان معامل الدقة في الحقيقة أكبر من هذه القيمة أو كم هو أكبر منه.

3) معامل ألفا هو متوسط المعاملات الممكنة للتجزئة النصفية المحسوبة بطريقة رولون. ولكن بشكل آخر، فإن كان معامل الثبات محسوب بالتجزئة النصفية العشوائية لفقرات الاختبار وحسب معامل رولون فإن معامل ألفا هو القيمة المتوقعة لهذا التقدير.

4) وأحد المفاهيم الخاطئة لمعامل ألفا، هو أن القيمة العالية نسبياً لهذا المعامل تشير إلى أن الاختبار أحادي البُعد (وهو أن الأداء على هذه الفقرات يمكن تفسيره بوساطة عامل واحد)، وذلك لأن معامل ألفا دالة للتباين المشترك بين الفقرات، ويمكن أن يكون التباين المشترك الكبير بين الفقرات نتيجة لأكثر من عامل أو بُعد واحد، ولا يجب تفسير معامل ألفا على أنه مقياس لأحادية البعد. فعلى سبيل المثال يمكن تفسير الدرجات على فقرات اختبار مقالي في العلوم الاجتماعية من خلال قدرات المفحوصين الكتابية، ومن خلال معرفتهم بالمحتوى، لذلك لا يمكن اعتبار الاختبار أحادي البعد (يقيس سمة واحدة)، ولأن فقرات الاختبار جميعها تتأثر بهاتين القدرتين فإن قيمة معامل ألفا تكون عالية لذا يمكن تفسير معامل ألفا على أنه الحد الأدنى من نسبة التباين في درجات الاختبار المفسر بالعوامل المشتركة المؤثرة في الأداء على فقرات الاختبار. وقدم مكدونالد (Mc Donald, 1981) مناقشة أكثر حداثة لهذه القضية عرض خلاله أمثلة عديدة مع براهين متعلقة بهذه النقطة.

وأخيراً: يجب ملاحظة أن معامل ألفا قابل للتطبيق على وجه العموم في أي موقف يحسب فيه ثبات المركب. ويمكن تطبيقه على سبيل المثال لتقدير ثبات الدرجة الكلية المعتمد على مجموع درجات اختبارات فرعية عديدة. والاستخدام الأكثر شيوعاً لمعامل ألفا هو اعتبار كل فقرة اختبار فرعي، ويمكن اعتبار الحالة الخاصة التي تكون فيها الدرجة الكلية مركب لدرجات نصفي الاختبار (مثل تلك التي تستخدم في التجزئة النصفية). ففي هذه الحالة وبوجود عنصرين فقط نرمز للمعامل بالرمز ($\alpha 2$) وهو مشابه لتقدير الثبات المحسوب بطريقة رولون أو جوتمان. لاحظ أن النتائج الناجمة عن استخدام ($\alpha 2$) تعتمد مكونات نصفي الاختبار، وقيمته لا تساوي بالضرورة قيمة معامل (α) الناتج عن معالجة كل فقرة على أنها عنصر مستقل في درجة الاختبار المركب.

ثبات تقديرات المقدرين: Inter rater Reliability

تستخدم في بعض أدوات القياس مجموعة واحدة من الفقرات (مثل قائمة السلوكات في استبانة الأداء)، ولكن تم جمع ملاحظات أو بيانات عديدة عن كل مفحوص من قبل أثنين أو أكثر من الملاحظين الذين يقيمون قائمة المقدرين. وفي هذا الموقف يكون المهم تجانس الملاحظات أو البيانات عبر المقدرين المختلفين. ولغاية اليوم فإن الأسلوب المفيد في حساب معامل ثبات المقدرين هو تطبيق معامل إمكانية التعميم المبين في الفصل الثامن، بالإضافة إلى معاملات أخرى للتوافق (مثل نسبة الإتفاق عبر المقدرين المؤشرة على فقرات بعينها أو مجموعة فقرات) تم مراجعتها من قبل فريك وسميل (Frick & Semmel, 1978) وهذه المعاملات تعد تشخيصية وتختلف مفاهيمياً عن تقديرات الثبات ولا يجب اعتبارها بدائل لتقديرات الثبات في وصف أدوات الملاحظة.

العوامل المؤثرة في معاملات الثبات:

تجانس المجموعة:

من الواضح أن قيمة معامل الثبات يعتمد على تباين الأفراد في درجاتهم الحقيقية والدرجات الخطأ. لذا فإن تجانس مجموعة المفحوصين يعد مهماً في تطوير الاختبار وفي اختيار الاختبار. افترض أن اختباراً طُوِّر لقياس القلق الرياضي، فإذا طبق هذا الاختبار على طلاب صف منتقى في الرياضيات، فمثل هؤلاء الطلاب من المحتمل أن يظهروا مستويات متشابهة ومنخفضة في القلق الرياضي، لذا فإن تباين الدرجات الحقيقية سيكون منخفضاً وكذا معامل الثبات. وإذا طبق الاختبار نفسه على مجموعة مساوية لعدد الطلبة ومؤلفة من مقطع عرضي من طلبة المرحلة الثانوية، وهنا تكون الدرجات الحقيقية أكثر تبايناً، وعلى افتراض ثبات تباين الخطأ العشوائي للمجموعتين المتساويتين في العدد فإن معامل ثبات المجموعة الثانية يجب أن يكون أكبر لأن تباين درجاتهم الحقيقية يمثل نسبة أكبر من تباين الدرجات الملاحظة. ويبين جدول (7-5) مثال افتراضي لهذا الموقف الذي يوضح أن الاختبار ليس ثابت (not reliable) أو غير ثابت (Unreliable) فضلاً عن أن الثبات هوخاصية لدرجات الاختبار ولمجموعة معينة من المفحوصين لذلك يحتاج مستخدم الاختبار لتحديد ما إذا كانت تقديرات الثبات الموجودة في أدلة الاختبارات تعتمد على عينة مشابهة في تركيبها وتنوعها للمجموعة التي سيطبق عليها الاختبار. فإن كانت عينة ناشر الاختبار غير متجانسة على السمة المقيسة بدرجة كبيرة فإن هذا يؤدي حتماً إلى خفض معامل الثبات عند تطبيق هذا الاختبار على عينة أكثر تجانساً وعلى وجه التحديد يمكن القول بأن الاختبار الصعب

جداً أو السهل جداً لفئة المفحوصين يؤدي إلى خفض مدى الدرجات، وكذا سيكون تباين الدرجات الحقيقية.

وقد قدم جوليكسن (Gulliksen, 1950) خلاصة تاريخية للجدل السيكومتري لعدم تجانس المجموعة وأثرها على ثبات الاختبار، وقدم ماغنسون (Magnusson, 1967, P.75) الصيغة الآتية للتنبؤ بتغير الثبات بتغير تباين العينة:

$$\rho_{vv'} = \frac{\sigma_x^2 (1- P_{xx1})}{\sigma_v^2} - 1 \quad (7 - 7) \text{ ....}$$

حيث يمثل σ_v^2 تباين العينة الجديدة

و σ_x^2 تباين العينة الأصلية

و $\rho_{xx'}$ تقدير ثبات العينة الأصلية

و $\rho_{vv'}$ تقدير الثبات المتنبأ به للعينة الجديدة

ومن المهم ملاحظة أن هذه المعادلة تفترض تساوي تباينات خطأ كلا المجموعتين، وأن التغير في الدرجات الملاحظة يعود إلى الفروقات في توزيع الدرجات الحقيقية للمجموعة، وحيث أن مستخدم الاختبار لا يمكنه الجزم بتحقيق هذا الافتراض عندما تكون العينة المستخدمة مختلفة بشكل ملحوظ عن العينة الأصلية، ومن الأنسب هنا التحقق من ثبات العينة الجديدة.

جدول (7-5): معاملات ثبات مجموعتين افتراضيتين تختلف في تباينات درجاتها الحقيقية

	المجموعة 1	المجموعة 2
تباين الدرجة الحقيقية	20	60
	10	
تباين الدرجة الخطأ	10	10
تباين الدرجة الملاحظة	30	70
معامل الثبات	67	0.85

إن اعتماد الثبات التقليدي على تباين الدرجة الحقيقية يتضمن أن معامل الثبات كما هو في التعريف له أهمية محدودة في تقويم نوعية المعلومات المستمدة من الاختبار لاستخدامها

في الغربلة أو الإنتقاء. وفي مثل هذه الحالات يهتم الفاحص فقط فيما إذا كانت درجة المفحوص أعلى أو أدنى من درجة قطع معينة. وقيم تباينات الدرجات الحقيقية والملاحظة (والنسبة بينهما) له صلة وثيقة أقل بعملية قياس الثبات، وهذا للاختبارات المستخدمة بالطريقة التي ستناقش في الفصل التاسع.

حدود الزمن Time Limit

عندما يكون وقت الاختبار محدداً بدقة بمعنى أن بعض المفحوصين أنهوا الاختبار وآخرين لم ينهوه، فإن مدى عمل المفحوص يؤثر بانتظام في أدائه على صيغ الاختبار جميعها. لذا فإن مدى عمل كل مفحوص يصبح جزءاً من تباين الدرجة الحقيقية. وفي بعض الاختبارات (مثل اختبارات الاستعداد الكتابية، واختبارات الحساب الرياضي للبالغين)، فقد يكون هدف الفاحص تقويم القدرة على أداء المهام بسرعة. وفي أنواع أخرى من الاختبارات قد يكون وقت الاختبار غير ضروري بالنسبة للسمة المقيسة. وفي هذا النوع من الاختبارات فإن وقت الاختبار يكون طويلاً بما فيه الكفاية ليسمح لمعظم الطلبة أو جميعهم بإنهاء الاختبار. عدا ذلك فإن تقدير الثبات يكون مبالغاً فية بسبب الاتساق في الأداء الناجم عن الوقت المحدد للاختبار وذلك عندما يكون الفاحص مهتماً بدرجة الاتساق في الأداء الاختباري الذي يجب ملاحظته في الموقف الذي ينهي فيه المفحوصين جميعهم الاختبار. وستجد في مؤلف لورد ونوفيك (Lord & Novick, 1968) تفصيلات أكثر لاختبارات السرعة.

ويكون التقدير مبالغ فيه بدرجة أكبر للثبات في اختبارات السرعة خاصة عند استخدام التجزئة بأسلوب اختبار الفقرات ذات الأرقام الفردية والزوجية. ومن الواضح أن الأداء على الفقرات خارج إطار الوقت المحدد للاختبار على الفقرات المتبقية الفردية والزوجية يكون اتساقها تاماً بغض النظر عما إذا كانت الفقرات متجانسة في محتواها. والأساليب الأخرى للاتساق تؤدي أيضاً إلى تقديرات مبالغ فيها للثبات للسبب نفسه عندما يكون الاختبار اختبار سرعة كبيرة، لذا فقد يكون استخدام طريقة الاختبار- الإعادة أو طريقة الصيغ المتكافئة أكثر ملائمة لتقدير ثبات اختبارات السرعة. وأسلوب آخر للصيغ المتكافئة هو تقسيم فقرات الاختبار إلى نصفين منفصلين وتطبيق كل نصف بصورة مستقلة وضمن الوقت الخاص به. ويحسب ثبات الاختبار بأكمله فيما بعد باستخدم معادلة إما سبيرمان- براون أو جوتمان. وبغض النظر عن الطريقة المستخدمة، فإن تقدير ثبات اختبارات السرعة يجب تفسيرها بحذر عندما تتطلب المهام الاختبارية أكثر من قدرة أداء مهام بسيطة بسرعة كبيرة.

طول الاختبار:

أحد مظاهر الاختبار التي تؤثر بالتأكيد على تباين الدرجة الحقيقية والملاحظة هو طول الاختبار. وهذا يتوضح فيما لو نظرنا إلى موقف يستخدم فيه الفاحص اختبار مؤلف من فقرة واحدة وآخر مؤلف من عشر فقرات (تعتمد جميعها على المحتوى نفسه). ومن الواضح أننا نضع ثقة أكبر على الدرجة الناتجة عن الاختبار الأطول، وقد اشتقت العلاقة بين طول الاختبار والثبات في الفصل السادس من خلال صيغة نبوءة سبيرمان– براون:

$$\frac{k\,P_{jj}^{11}}{1+(k-1)P_{jj}^{1}} = \rho_{xx}^{\backslash}$$

حيث ترمز ρ_{jj}^{1} إلى ثبات اختبار فرعي منفرد

و Ke إلى عدد الاختبارات الفرعية في المركب

$\rho_{xx}^{\backslash}$ إلى ثبات المركب

فإن كان ثبات المجموعة الفرعية المنفردة تساوي (.75)، فإن ثبات الاختبار المؤلف من خمس فقرات فرعية متوازية تحسب على النحو الآتي:

$$0.94 = \frac{(0.75)\,5}{(0.75)\,4+1} = \rho_{xx}^{\backslash}$$

كيف يمكن عزو هذا إلى طول الاختبار؟ بصورة أساسية نفترض أن الاختبار الذي بمتناول أيدينا هو الاختبار الفرعي j، وثباته المحسوب هو P_{jj}^{1} وباستخدام أية طريقة مناسبة، نفترض بعدها أن الاختبار الجديد مركب ومؤلف من اختبارات فرعية عددها j على أنه عنصر واحد، ثم إن k هي العامل الذي سيتضاعف إليه طول الاختبار. وعلى سبيل المثال إن كان الاختبار j مؤلفاً من (50) فقرة، والاختبار بالطول الجديد مؤلف من (150) فقرة، فإن الاختبار الجديد يصوّر على أنه مركب من ثلاث اختبارات فرعية كل منها مؤلف من (50) فقرة، أو أن k = 3. وعلاوة على ذلك فإن k قد لا تكون قيمتها عدد صحيح دائماً أو أكبر من (1). فعلى سبيل المثال افترض أن الاختبار الحالي مؤلف من (100) فقرة، ويرغب مطور الاختبار معرفة مدى تأثير الثبات عند تقصير الاختبار إلى (75) فقرة. ولأن k هي العامل الذي يجب أن يضرب بطول الاختبار وصولاً إلى الطول الجديد (100 (k) = 75) أو أن (k = .75). فإن كانت . $0.60=\rho_{jj}^{1}$ فإن ثبات الاختبار الجديد القصير يكون:

$$0.53 = \frac{(0.60\ )\ x\ 0.75}{0.6\ (1\text{-}0.75)\ +1} = \rho_{xx}{}^{\backslash}$$

وعندما يزداد طول الاختبار فإن قيمة k تصبح أكبر من (1)، وعندما يقل طول الاختبار فإن قيمة k تصبح أقل من (1). ويجب ملاحظة أن الزيادة في ثبات الاختبار الناتجة عن زيادة طول الاختبار تتبع قانون تناقص الغلة (وينص هذا القانون على أن زيادة العمل أو رأس المال إلى أبعد من نقطة معينة لا يترتب عليه زيادة مناسبة في الانتاج). وتطبيق هذا القانون هنا يبين أن مضاعفة طول اختبار ثباته (.60) يزيد ثباته إلى (.75)، وسيزيد ثلاثة أضعاف الطول ثبات الاختبار إلى (.81)، ولكن مضاعفة الطول إلى خمسة أضعاف طوله الأصلي سيزيد الثبات إلى (.88) فقط. لذلك وعند نقطة معينة فإن الزيادة القليلة في الثبات الناتجة إضافة فقرات من المحتمل أن لا تغطي كلفة كتابة الفقرات الاضافية والوقت المستهلك في تطبيق الاختبار. وعلاوة على ذلك فإن تصحيح سبيرمان-براون هو انعكاس دقيق للثبات فقط إذا كانت الفقرات المضافة أو المحذوفة موازية في المحتوى والصعوبة لفقرات الاختبار الأصلي من (x). لذا فإن أي قيمة متنبأ بها للدرجة الحقيقية لفرد معين ستكون أقرب إلى وسط المجموعة من الدرجة الملاحظة الأصلية. إضافة إلى أن معامل الثبات الأقل يؤدي إلى أن الدرجة الحقيقية المتنبأ بها أقرب إلى وسط المجموعة، وعندما تكون قيمة μ_{xx} = صفر لأي فرد فإن قيمة $T^{\backslash}$ تقترب من قيمة x.

ولأن الدرجات الحقيقية المقدرة يمكن حسابها بسهولة فإن مستخدم الاختبار العرضي قد يدعي استخدامها على أساس أن الدرجة الحقيقية المقدرة ينتج عنها قرارات أكثر ملائمة عن المفحوصين. وليس من الضروري أن يكون الموقف بهذا الشكل، وكقانون عام لا توجد ميزة لحساب الدرجات الحقيقية لمجموعة واحدة من المفحوصين التي ستتم مقارنتها على أساس معياري- المرجع. ومستخدم الاختبار الذي سيحسب الدرجات الخام والدرجات الحقيقية المقدرة للطلبة سيجد أن وسط كلا التوزيعين هو نفسه، إضافة إلى أن ترتيب المفحوصين سيكون هو نفسه لكلا التوزيعين. وفقط تختلف الإنحرافات المعيارية لكلا التوزيعين حيث يكون الانحراف المعياري للدرجات الحقيقية هو الأقل.

تقدير الدرجات الحقيقية:

مع أن قيمة الدرجة الحقيقية لمفحوص معين لا يمكن تحديدها بدقة، لكن يمكن الحصول على تقدير لها من خلال معادلة الإنحدار العامة للتنبؤ بقيمة Y من خلال قيمة معروفة X، تذكر أن:

$$\rho_{xy}\ \frac{\sigma_y}{\sigma_x{}^2}\ (x - \mu_x) + \mu_Y = Y_1$$

وبالتعويض بـ : ρ_{XT} لـ ρ_{Xy} و $T^\backslash$ لـ $y^\backslash$

نحصل على:

$$\rho_{XT} \frac{\sigma_T}{\sigma_x} (x - \mu_x) + \mu_T = T^\backslash$$

ولأن

و $\rho_{XT} = \frac{\sigma_T}{\sigma_x}$ ،

و $\rho_{XT}^2 = \rho_{XX^\backslash}$

و $\mu_T = \mu_x$ ، بالتعويض المناسب نصل إلى:

$$\rho_{XX^\backslash} (X - \mu_x) + \mu_T = T^\backslash \quad \text{............ } (9 - 7)$$

وإن أعدنا كتابة المعادلة (7-9) بدلالة الدرجات المنحرفة حيث أن:

$$t_1 = T_1 - \mu_x$$

و $x = X - \mu_x$ نصل إلى:

$$t^\backslash = \rho_{XX^\backslash}(X)$$

وحيث أن أي تقدير لـ$\rho_{XX^\backslash}(X)$ يعتمد على بيانات حقيقية يكون أقل من (1)، لذلك فإن $t^\backslash$ ستكون أقل.

وفي المواقف التي سيتم فيها تحديد مواقع مختلفة للمفحوصين اعتماداً على تواجدهم أعلى أو أدنى من نقطة معينة، فإن استخدام الدرجات الحقيقية المقدرة سيؤدي إلى نتائج مبهمة يجب التحقق منها بجدية قبل التطبيق. خذ مثلاً الحالة التي سيوزع فيها الطلبة على برامج علاجية أو خاصة اعتماداً على نتائج الاختبار. فالطلبة الذين درجاتهم أقل من (80) سيتم وضعهم في صفوف المعاقين عقلياً القابلين للتعلم، والطلبة الذين درجاتهم أعلى من (130) سيتم وضعهم في صفوف الموهوبين.

افترض أن جوزيف حصل على درجة (79) في اختبار الذكاء وكارين حصلت على درجة (132)، سيعامل كل منهم عل أنه حالة خاصة وتقدم له خدمات خاصة، ولكن باستخدام الدرجات الحقيقية المقدرة عندما يكون وسط المجموعة (100)، و $P_{xx1} = .90$، فإن درجة جوزيف الحقيقية المقدرة تكون:

$$T_J = .9 \ (79 - 100) + 100 = 81.1$$

ودرجة كارين الحقيقية المقدرة تكون:

$$128.8 = 100 + (100 - 132)\ .9 = T_k$$

ومنها نرى أنه لا يحتاج أي منهما إلى برامج خاصة. بالمقابل أنظر كيف ستتغير النتائج فيما لو استخدمت وسط المجموعة الفرعية (المعاقين عقلياً) بدلاً من وسط المجموعة الكلي. فإن كان وسط هذه المجموعة (65)، فإن درجة جوزيف الحقيقية المقدرة ستكون:

$$77.6 = 65 + (65 - 79)\ .90 = T_j$$

وبالاعتماد على أسس الدرجة الحقيقية المقدرة سيتم تصنيف جوزيف على أنه معاق عقلياً. ويوضح هذا المثال أهمية تحديد المجموعة المعيارية المناسبة فيما لو استخدمت الدرجات الحقيقية المقدرة لاتخاذ قرارات تتعلق بالمفحوصين.

وكقاعدة عامة، من الأفضل استخدام وسط المجموعة الكلية بدلاً من أوساط المجموعات الفرعية التي تعتمد على مجموعات صغيرة نسبياً وبالتالي تكون غير مستقرة. فهو تطبيق معرض للتساؤل، إن تولد مجموعات فرعية اعتماداً على درجات الاختبار ومن ثم استخدام أوساط المجموعات الفرعية في تقدير الدرجات الحقيقية على الاختبار نفسه أو اختبار مشابه له (كما تم توضيح ذلك في المثال السابق). إن استخدام أوساط المجموعات الفرعية في تقدير الدرجات الحقيقية تكون مُسوّغة عندما تكون المجموعات الفرعية معتمدة على متغيرات ديموغرافية طبيعية أو تعليمية تدريسية لا علاقة لها بالاختبار. على سبيل المثال عند تقدير الدرجات الحقيقية للسود والبيض والاسبان نأمل استخدام وسط المجموعة العرقية الفرعية لا وسط المجموعة الكلية.

مع ذلك ففي معظم المواقف يكون تقدير الدرجة الحقيقية غير ضروري، وسنبين باختصار أن هنالك أنواع من البحوث أو التقويم حيث يفضل اعتماد التقدير للدرجة الحقيقية. والمشكلة التي تظهر في بعض الدراسات هو مقارنة معادلات انحدار مجموعتين أو أكثر التي اوساطها غير متساوية على المتنبيء. وقد تظهر هذه المشكلة عند استخدام تحليل التباين المشترك في تصميم المجموعات غير المتكافئة (Campbell & Stanely, 1963)، وكذا في دراسة تحيز الفقرات. ان استخدام الدرجات الملاحظة يمكن أن يخلق مجموعة فروق في معادلات الانحدار التي قد لا تظهر فيما لو توافرت الدرجات الحقيقية. وقد بين هنتر وكوهين (Hunter & Cohen, 1974)، إن استخدام الدرجات الحقيقية المقدرة قد يتغلب على هذه المشكلة في الانحدار الخطي ويقللها في تحليل الانحدار غير الخطي.

ثبات الفروق في الدرجات:

هنالك مواقف بحثية وتقويمية وتحليل تشخيصي يطبق فيها اختباران، والمتغير الذي يهتم

به هو الفرق بين الدرجات على الاختبارين، وفيما يأتي مثالان واسعا الانتشار ينطبق عليها هذا:

1) عندما يهدف المقوِّم تحديد الكسب في الأداء لكل مفحوص عبر الزمن وباستخدام الاختبار نفسه.

2) محلل نفسي متخصص في صعوبات التعلم مهتم في الفجوة في أداء المفحوصين على اختبارين أو على اختبارين فرعيين (مثل درجات الاختبارات الفرعية في تطور اللغة، درجات الاختبارات الفرعية في إنتاج اللغة).

ففي كلا الحالتين المتغير الذي نهتم به هو الفرق في الدرجات، والذي يُحدد على النحو الآتي:

$$X - Y = D$$

حيث تمثل x الدرجة على القياس الأول

و y الدرجة على القياس الثاني، فإن كانت القرارات عن الأفراد ستُعتمد بناءاً على فروق الدرجات فيجب حساب ثبات الفروق في الدرجات.

ويمكن اشتقاق الصيغة التقليدية لحساب ثبات فروق الدرجات على النحو الآتي:

في البدء يمكننا تمييز الفروق بين الدرجات الحقيقية على المتغيرين x و y كما يأتي:

$$T_x - T_y = T_D$$

وحيث أن كل من D و TD عبارة عن مركبات خطية للدرجات الأخرى، فإن تباينهما يمكن التعبير عنه على النحو:

$$\sigma^2_x + \sigma^2_y - 2\sigma_{xy} = \sigma_D^2$$

و $$\sigma^2_{Tx} + \sigma^2_{Ty} - 2\sigma_{xy} = \sigma_{TD}^2$$

ومن التعريف الأساسي للثبات نعرف أيضاً أن:

$$\rho_{xx'} \ \sigma^2_x = \sigma^2_{Tx}$$

وأن $$\rho_{xy'} \ \sigma^2_y = \sigma^2_{Ty}$$

كذلك، فقد عرفنا من الفصل السادس أن:

$$\sigma_{TxTy} = \sigma_{xy}$$

وبإجراء التعويضات الثلاث يمكن التعبير عن σ^2_{TD} من خلال الدرجات الملاحظة على النحو الآتي:

$$\rho_{XX'}\sigma^2_x + \rho_{XX'}\sigma^2_y - 2\sigma_{xy} = \sigma^2_{TD}$$

ومن خلال التعريف النظري لثبات الفرق بين الدرجات:

$$\rho_{DD'} = \frac{\sigma^2_{TD}}{\sigma^2_D} \quad \text{.............. (10 - 7)}$$

وبإجراء التعويض المناسب لكل من : σ^2_D و σ^2_{TD}

والتعبير عن σ_{xy} بدلالة $\rho_{Xy}\,\sigma_x\sigma_y$ نصل إلى:

$$\rho_{DD'} = \frac{\rho_{XXx}\,\sigma^2_x + \rho_{yyx}\,\sigma^2_y - 2\,\rho_{xy}\,\sigma_x\sigma_y}{\sigma^2_x + \sigma^2_y - 2\,\rho_{xy}\,\sigma_x\sigma_y} \quad \text{.............. (11 - 7)}$$

ولنرى كيف نطبق هذه الصيغة عملياً، افترض أن الباحث خطط لدراسة الأطفال الذين يظهرون فجوة بين قدرتي إنتاج اللغة وعملياتها. ويبين الجدول (7-6) بيانات افتراضية لهذا المثال، فإن وجد الباحث ارتباطاً بين إنتاج اللغة، وعملياتها ($P_{xy} = .70$)، فإن ثبات الفرق بين درجات الاختبارات الفرعية سيكون:

جدول (7-6): حسابات افتراضية للوسط، الانحراف المعياري، الثبات للاختبارات الفرعية في إنتاج اللغة وعملياتها للأطفال.

	اختبار إنتاج اللغة	اختبار عمليات اللغة
الوسط	32	40
الانحراف المعياري	6	8
الثبات	.75	.80

$$\rho_{DD'} = \frac{(8)\,(6)\,(0.7)\,2 - (64)\,(0.8) + (36)\,(0.75)}{(8)\,(6)\,(0.7)\,2 + 64 + 36}$$

$$= 0.34$$

من جهة أخرى، إن كان الارتباط بين الاختبارات الفرعية ($P_{Xy} = .30$) فإن ثبات الفرق بين الدرجات سيكون أعلى.

$$0.69 = \frac{(8)\,(6)\,(0.3)\,2 - (0.64)\,(0.8) + (0.36)\,(0.75)}{(8)\,(6)\,(0.3)\,2 - 64 + 36} = \rho_{DD1}$$

لذا نرى أن ثبات الفرق بين الدرجات يكون أكبر عند اختيار اختبارين أو مقياسين ثباتهما عالٍ، ولكن الارتباط بين القياسين منخفض، ومع ذلك توجد استثناءات في بعض المواقف لهذا القانون، فعلى سبيل المثال: أجرى زيمرمان وويليامز (Zimmerman & Williams, 1982) تطبيقات لتجميعات شروط كانت فروق الدرجات فيها ثابتة مع وجود ارتباط لها مع درجات الاختبارات القبلية.

استخدام تقديرات الخطأ في تفسير الدرجات:

قد تفيد تقديرات الثبات في المقارنة بين درجات اختبارين أو أكثر، ولكن عند تفسير درجة المفحوص كفرد فإن معامل الخطأ المتوقع في درجة الاختبار تكون أكثر أهمية. وتاريخياً فإن ثلاثة أنواع من الأخطاء تم تحديدها بالأساس من قبل كيلي (Kelley, 1927) وتعد ملائمة في تفسير درجة المفحوص، وهي:

1) الفرق بين الدرجة الحقيقية والدرجة الملاحظة للمفحوص في موقف اختباري معين.

2) الفرق بين درجة الفرد الملاحظة على اختبار والدرجة المتوقعة على صيغة اختبارية مكافئة.

3) الفرق بين درجة الفرد الحقيقية ودرجته الحقيقية المقدرة.

فإن كان اهتمام مستخدم الاختبار بالخطأ من النوع الأول، فسنستخدم الخطأ المعياري للقياس. وكما لوحظ في الفصل السادس يمكننا تصور الخطأ العياري للقياس على أنه الانحراف المعياري للفرق بين درجة المفحوص الحقيقية ودرجته الملاحظة لعدد لا نهائي من المواقف الاختبارية المتكررة. وكتقدير للخطأ المعياري للقياس يمكن تحديده باستخدام الصيغة:

$$\sigma_x \sqrt{1-\hat{\rho}_{xx'}} = \hat{\sigma}_E$$

وتمثل σ_x الانحراف المعياري للدرجات الملاحظة للمجموعة كلها و $\rho_{xx'}$ تقدير الثبات، ويبين الجدول (7-7) الحساب التوضيحي والتفسير الاحصائي لفحوص درجته الملاحظة (30).

ويتطلب تفسير الدرجات باستخدام الخطأ المعياري تأسيس فترة ثقة حول الدرجة الملاحظة نتوقع أن تقع الدرجة ضمنها. وسنستخدم في المناقشة التالية الخطأ المعياري للقياس، وتطبق هذه النقاط بالطريقة نفسها على الأخطاء من النوعين الآخرين (الثاني والثالث). وتذكر من الفصل السادس أن الصيغة العامة لفترة الثقة باستخدام الخطأ المعياري للقياس هي: $(x \mp 1\sigma_E)$ إن أردنا أن تكون درجة الثقة (68%) لأن تقع الدرجة الحقيقية ضمن الفترة، أما إن أردنا أن تكون درجة الثقة بنسبة (95%) فإن الفترة تكون $(x \mp 1.96\,\hat{\sigma}_E)$ أو حول $(x \mp 2\sigma_E)$

ويوضح جدول (7-7) استخدام هذه الصيغة والقيم الناتجة عن تطبيقها على الدرجة الملاحظة لمفحوص على اختبار معين. وكذلك يبين الجدول التفسير باستخدام فترة الثقة هذه

جدول (7-7): الصيغ والحسابات والتفسيرات لدرجات الثقة باستخدام تقديرات الخطأ المعياري للمفحوص نفسه

$$[\,\hat{\mu} = 26 \,،\; \hat{\sigma}_{1x} = \hat{\sigma}_{2x} = 10 \,،\; \hat{\rho}_{xx'} = 0.75\,]$$

الخطأ المعياري	الصيغة	فترة ثقة 68%	فترة ثقة 95%	تفسير فترة الثقة 95% لمفحوص درجته الملاحظة (30)
الخطأ المعياري للقياس	$\sigma_E = \sigma_x\sqrt{1-\rho_{xx'}}$ $\sigma_E = 10\sqrt{1-0.75}$ $\sigma_E = 5$	30 ± 5	30 ± 10	نثق بنسبة 95% بأن درجة شارلوت الحقيقية تقع في الفترة 20-40
الخطأ المعياري للتقدير	$\sigma_{1.2} = \sigma_x\sqrt{1-\rho_{xx'}^2}$ $\hat{\sigma}_{1.2} = 10\sqrt{1-(0.75)^2}$ $\hat{\sigma}_{1.2} = 6.6$	30 ± 6.6	30 ± 13.2	نثق بنسبة 95% أن شارلوت لو أعيد اختباره على اختبار موازي فإن درجته على الاختبار الثاني تقع في الفترة (16.8 - 43.2)

للمفحوص نفسه. ومن المهم تمييز أن (95%) من المفحوصين الذين تقدموا للاختبار تقع درجاتهم الحقيقية ضمن فترة الثقة المحسوبة بهذه الصيغة. لذلك ولمفحوص معين مثل شارلوت فإن هنالك فرصة 95% بأن فترة ثقة حول درجة الملاحظة تتضمن درجته الحقيقية. ومع ذلك لا يمكننا القول بأن كل مفحوص وبثقة مطلقة بأن الفترة تتضمن درجته الحقيقية (فقد يكون شارلوت أحد الـ 5% الذين تقع درجاتهم الحقيقية خارج هذه الفترة). إضافة إلى أنه ليس صحيحاً لجميع المفحوصين الذين حصلوا على درجة ملاحظة معينة (مثل 30) بأن تكون 95% من درجاتهم الحقيقية تقع ضمن فترة الثقة المحددة.

في العديد من المواقف في تفسير درجات الاختبار مع الطلبة نفسهم أو والديهم أو المرشدين لا يكون هنالك جدوى من مناقشة الدرجات الحقيقية معهم. ويبدو أنه أكثر جدوى ونفعاً مناقشة الدرجة التي سيحصل عليها المفحوص فيما لو تقدم لصيغة موازية أو بديلة للاختبار نفسه. وفي مثل هذه الحالة يستخدم الخطأ من النوع الثاني والخطأ المعياري للتقدير للتنبؤ بدرجة المفحوص على الصيغة 2 من درجته الملاحظة على الصيغة 1 تكون على النحو الآتي:

$$\sigma_{1.2} = \hat{\sigma}_{X2}\sqrt{1 - P_{XX'}^{2}}$$

وثانية، فإن الحسابات والتفسير الاحصائي مبين في الجدول (7-7). ومن الملفت للنظر أن فترة الثقة الناتجة عن استخدام الخطأ المعياري للتقدير تكون أكبر من فترة الثقة المحسوبة بناءاً على الخطأ المعياري للقياس. وهذا منطقي، فإذا تذكرنا أن الخطأ المعياري للقياس على الاختبارين الأصلي والصيغة البديلة سيؤثر على فترة الثقة الثانية.

تدوين بيانات الثبات:

يقع على عاتق مطور الاختبار الالتزام بمسؤوليته ليس حساب ثبات درجات الاختبار في ظروف مختلفة فحسب. بل أيصال هذه المعلومات إلى مستخدمي الاختبارات على هيئة معلومات. وقد قدمت الرابطة المسماة المعايير للاختبارات التربوية والنفسية (AERA-APA NCME, 1985) وسابقتها المعايير للاختبارات التربوية والنفسية (APA, AERA NCME, 1974) بإرشادات مهمة للمعايير التي يجب أن يراعيها مطور الاختبار. ومن هذه الإرشادات النقاط الآتية والتي تناسب الفصلين السادس والسابع وهي:

1) يجب كتابة نتائج دراسات ثبات مختلفة لتؤخذ بعين الاعتبار مصادر الخطأ المختلفة

والأكثر ملائمة لاستخدام الدرجات، ومع ذلك فمن المناسب كتابة مؤشرات الاتساق الداخلي لغير اختبارات السرعة، ولا تعد مثل هذه الأدلة بديلاً عن المعلومات حول استقرار درجات الاختبار.

2) الأخطاء المعيارية للقياس ولتجمعات الدرجات لفترات الثقة المختلفة يجب أن ترافق كل تقدير للثبات. ويجب أن تكتب الأخطاء المعيارية بوحدات التدريج لكل الدرجات المعيارية المتوافرة.

3) يجب كتابة تقديرات الثبات والخطأ المعياري لدرجات الاختبارات الفرعية كما هو الحال لدرجات الاختبار الكلي.

4) يجب وصف الطرائق والعينات المستخدمة في دراسات الثبات وصفاً كافياً ليسمح للمستخدمين تحديد التشابه بين ظروف دراسة الثبات والموقف الخاص بمستخدم الاختبار.

5) عند استخدام الاختبار لمجتمع مفحوصين معين له التوزيع الطبيعي (مثل طلبة مستوى صفي معين أو ذوي الإعاقات الخاصة) يجب أن تدون تقديرات الثبات وأخطاء القياس المعيارية لها منفصلة.

6) وعندما تستخدم درجات الاختبار بصورة أولية في وصف أو مقارنة أداء مجموعة معينة فيجب تدوين الثبات وخطأ القياس المعياري للملاحظات المتجمعة.

7) إذا حسبت الأخطاء المعيارية للقياس باستخدام أنموذج معين مثل الإنموذج ثنائي الحد يجب الإشارة إلى ذلك، وعدا ذلك فإن مستخدمي الاختبارات يفترضون أن القيم المكتوبة هي قيمة خطأ القياس المعياري التقليدي.

الخلاصة:

من المهم لمطور الاختبار أن يحدد أنواع خطأ القياس الأكثر أهمية لمستخدمي الاختبار من أجل تصميم دراسات الثبات التي تسمح بتقييم أثر أخطاء القياس هذه على ثبات درجات الاختبار. ويمكن استخدام طرائق متنوعة لجمع وتحليل البيانات في دراسة الثبات. وعبر هذه الأساليب هنالك أساليب تتطلب تطبيقين منفصلين على مجموعة المفحوصين نفسها. وهذه الأساليب هي أسلوب الاختبار- الإعادة، وأسلوب الصيغ المتكافئة أو المتبادلة وأسلوب الاختبار- الاعادة باستخدام الصيغ المتبادلة. ونوع آخر من الدراسات تستخدم تطبيق

صيغة اختبارية واحدة على مجموعة المفحوصين، ويؤدي تحليل بيانات مثل هذه الدراسات إلى الحصول على معامل اتساق داخلي يزودنا بتقدير اتساق أداء المفحوصين عبر الفقرات في تطبيق واحد.

أحد أساليب تقييم الاتساق الداخلي هو التصحيح المنفصل لنصفي الاختبار لكل مفحوص. ويحسب معامل الارتباط بينهما ومن ثم يصحح باستخدام نبوءة سبيرمان- براون أو أن يحسب الفرق بين درجات نصفي الاختبار ومن ثم يحسب الثبات باستخدام طريقة رولون. ومحددات استخدام طرائق نصفي الاختبار هو إن طرائق التجزئة المختلفة ينتج عنها تقديرات ثبات مختلفة. وللتغلب على هذه المشكلة تحسب نسبة مجموع التباين المشترك للفقرات إلى تباين الدرجات الملاحظة الكلي. وهنالك ثلاث صيغ تعتمد هذا المبدأ وتؤدي جميعها إلى النتيجة نفسها وهي: كيودر- ريتشاردسون 20، وتحليل التباين لهويت، وكرونباخ الفا. ومعامل ألفا عبارة عن متوسط معاملات التجزئة جميعها الذي نحصل عليه فيما لو تمت تجزئة الاختبار إلى الأنصاف الممكنة جميعها، وحُسب الثبات باستخدام طريقة رولون. وعندما نفترض تساوي صعوبات الفقرات فإنه يمكن استخدام معادلة كيودر- ريتشاردسون 21 (KR21) بدلاً من معادلة كيودر- ريتشارسون 20 (KR20)، ويمكن حساب (KR21) من المتوسط والتباين وعدد فقرات الاختبار. وتتأثر تقديرات الثبات بعوامل عدة في الموقف الاختباري. فإن كانت عينة المفحوصين متجانسة بشكل كبير في السمة المقيسة فإن الثبات المحسوب يكون أقل مما لو كانت العينة غير متجانسة. وكذلك فإن الاختبار الأطول أكثر ثباتاً من الاختبارات الأقصر والمؤلفة من فقرات مشابهة. ويمكن حساب أثر تغيير طول الاختبار باستخدام نبوءة سبيرمان- براون. وتؤثر السرعة أيضاً (وقت الاختبار) على معامل الثبات.

وعند معرفة درجة المفحوص الملاحظة ومتوسط المجموعة ومعامل ثبات الاختبار فمن الممكن تقدير درجة المفحوص الحقيقية في الاختبار. وتكون الدرجات الحقيقية المحسوبة أقرب إلى متوسط المجموعة منها إلى الدرجات الخام.

وتكون فروق الدرجات أو الكسب في الدرجات بين الموقفين الاختباريين أقل ثباتاً من درجات أي من الموقفين الاختباريين عندما لا يكون هنالك ارتباط بين أخطاء القياس. ويمكن الحصول على ثبات فروق الدرجات فقط عند استخدام اختبارات ثباتها عالٍ في كلا التطبيقين أو الموقفين، وفي الوقت نفسه ارتباط

منخفض بين درجات الاختبارين.

وفي تفسير درجات الاختبار، توجد على الأقل ثلاثة أنواع من أخطاء القياس المناسبة هي: الخطأ المعياري للقياس، وهو مقياس للفجوة المحتملة بين درجتي المفحوص الحقيقية والملاحظة. والخطأ المعياري للتنبؤ، وهو مقياس للفجوة بين الدرجات الملاحظة على تطبيقين متتاليين للاختبار. والخطأ المعياري للتنبؤ بالدرجات الحقيقية، هو مقياس للفجوة بين تقديرات الدرجات الحقيقية المقدرة والدرجات الحقيقية الفعلية. ويمكن استخدام مقاييس الأخطاء المعيارية جميعها في تكوين فترة ثقة حول درجة المفحوص وذلك لتذكر مستخدمي الاختبارات بمدى الانحراف المتوقع في الدرجات.

وتزودنا معايير AE RA / APA/ NCME للاختبارات التربوية والنفسية بتعليمات وارشادات تفيد مطوري الاختبارات في تدوين ثبات درجات الاختبار. وكل مطور اختبار ملزم بمعرفة التعليمات واتباعها في اجراء دراسات الثبات وتدوين نتائجها.

التمارين:

1/ حدد نوع الثبات الأكثر ملائمة لكل موقف من المواقف الآتية:

أ- يرغب معلم التحقق من ثبات اختبار نهائي في الأحياء فقراته من نوع الاختيار من متعدد.

ب- يرغب عالم نفس ارشادي في تطوير استبانتين للاتجاهات واحدة لاستخدامها كاختبار قبلي والأخرى كاختبار بعدي في برنامج تربوي يستغرق يوم واحد.

جـ- طور عالم نفس اجتماعي أداة تقييم اتجاهات للمسجلين الجدد في الكلية، وأعطى كل فقرة نقطتين لاستجابة «نعم»، ونقطة واحدة لاستجابة «ممكن»، وصفر لاستجابة «لا»، وأجاب الجميع عن الاستبانة أثناء الأسبوع الارشادي.

د- صيغة اختبارية واحدة لاختبار تحصيلي يغطي العلوم الصحية بفقرات ثنائية التصحيح تغطي مجالات الصحة وضبط المرض والتكاثر. ويستفسر مطور الاختبار عما إذا كانت درجة كلية واحدة أفضل أم درجات المجالات الفرعية؟

هـ- طبق اختبار القبول لكلية دولية في فصل الربيع وبصيغ مختلفة استخدمت في موقفين. ويريد المشرفين في الكلية على اختبار القبول أدلة تؤشر على أن وقت التقدم للاختبار لا يؤثر في نتيجة الاختبار.

2/ طبقت استبانة اتجاهات ثنائية التصحيح على عشرة مستجيبين وكانت مصفوفة المفحوص- الاستجابة على النحو الآتي:

المفحوص	الفقرات								
	1	2	3	4	5	6	7	8	المجموع
أ	1	1	1	1	1	1	0	0	6
ب	1	1	1	0	0	1	0	0	4
جـ	1	1	0	1	0	0	0	0	3
د	0	0	0	0	1	0	1	1	3
هـ	1	1	1	1	1	1	1	1	8
و	1	1	1	1	1	0	0	0	5
ز	0	1	1	1	1	0	0	1	5
ح	1	1	1	1	1	1	1	0	7
ط	1	1	0	0	0	0	0	0	2
ي	1	1	1	1	1	0	1	1	7

أ) افترض أن الباحث يريد تقييم الاتساق الداخلي للاختبار باستخدام التجزئة النصفية، وذلك بتقسيم الاختبار إلى اختبارين فرعيين أحدهما بالفقرات الفردية والآخر بالفقرات الزوجية. احسب الدرجات التي حصل عليها كل مفحوص بهذه الطريقة.

ب) احسب معامل الارتباط بين هذين الاختبارين الفرعيين.

جـ) ما تقدير الثبات للاختبار الكلي.

د) بالاعتماد على الفحص البصري، هل تتوقع أن طريقة التجزئة باستخدام النصفين الأول والأخير للاختبار ينتج عنه تقدير أعلى للثبات أم أقل من ذلك الناتج عن التجزئة الفردية - الزوجية؟

هـ) أي طريقة للتجزئة ينتج عنها تقدير أكثر دقة لمعامل الدقة؟ عدّل استجابتك.

و) ما قيمة معامل كيودر–ريتشاردسون 20 لهذا الاختبار؟ ومعامل كيودر–ريتشاردسون 21؟

ز) فسر الاختلاف بين قيمتي KR20 و KR21؟

3/ يتضمن الجدول الآتي بيانات وصفية لاختبارين فرعيين لبطارية اختبارات معيارية، والذي يتضمن صيغتين بديلتين. تفحص هذه البيانات وأجب عن الأسئلة التالية:

	الحسابات الرياضية	حل المسألة الرياضي
الصيغة أ		
الوسط	22	25
الانحراف المعياري	8	12
عدد الفقرات	35	40
KR20	.83	.85
الصيغة ب		
الوسط	22	24
الانحراف المعياري	7	12
عدد الفقرات	35	40
KR20	.80	.85

أ) ما الأخطاء المعيارية للقياس لكلا الصيغتين؟ وهل الأفضل استخدام صيغة الاختبار ذات الخطأ المعياري الأقل؟

ب) إذا حصل توم على الدرجة 20 في اختبار الحسابات الرياضية على الصيغة أ، أسس فترة تتضمن درجته الحقيقية وبنسبة ثقة 68%. ما هي حدود هذه الفترة المحتملة لو أردنا أن تصبح نسبة الثقة 95%.

جـ) للصيغة الثانية في الحسابات الرياضية حصل جيل على الدرجة 10 وحصل هيريت على الدرجة 29. فما قيمة درجاتهم الحقيقية المقدرة.

د) فكر ناشر الاختبار في تقصير اختبار حل المشكلات الرياضية لتصبح خمس فقرات لكل صيغة. احسب كيف يتأثر تقدير الثبات.

هـ) في دراسة الثبات لم يكن الوقت كافياً لإكمال اختبار حل المشكلات من قبل جميع الطلبة، ما الأثر المحتمل لهذا على قيمة KR20.

و) افترض أن ناشر الاختبار دون ارتباط بقيمة (.80) بين اختبارين فرعيين من الصيغة أ وارتباط قيمته (.70) بين اختبارين فرعيين من الصيغة ب. واقترح الاستاذ استخدام درجات هذين الاختبارين في تحديد الطلبة الذين قد يستفيدوا من التعليم الإضافي لحقائق الرياضيات الأساسية. واقترح الاستاذ فحص الفجوة بين درجتي z لكل من الحساب الرياضي وحل المشكلات وللطلبة جميعهم. وللطلبة الذين كان فرق درجاتهم (الفجوة) سالب وبقيمة كبيرة تم إعطاءهم تعليم خاص. أي الصيغتين تؤدي إلى نتائج ثباتها أكبر؟ وما تقدير الثبات المترتب على فروق الدرجات هذه؟

4/ للاختبارات الموصوفة في الفقرة السابقة، يتضمن دليل الاختبار الجداول الآتية للبيانات الارتباطية. وبعد تطبيق الاختبار سأل عدة معلمين المشرف على الاختبار عن المعاملات التي يهتمون بها. أشر إلى أي المعاملات في المصفوفة تتضمن معلومات وجدها المشرف مناسبة لاهتمامات كل معلم.

	إعادة الاختبار بعد يوم	إعادة الاختبار بعد أسبوع
الصيغة أ		
الحسابات الرياضية	.95	.86
حل المسألة الرياضية	.87	.80
الصيغة ب		
الحسابات الرياضية	.91	.83
حل المسألة الرياضية	.89	.78

		الصيغة ب	
		الحسابات الرياضية	حل المسألة
الصيغة	الحسابات الرياضية	0.85	0.70
أ	حل المسألة	0.68	0.89

أ) اقترح جوين ليف فصله بيوم إجازة، وسألت إن كان ممكناً تطبيق الاختبار نفسه لطلبتها في اليوم التالي.

ب) قال السيد روبنسون أنه يتوقع أن الصيغة ب تتضمن أسئلة كثيرة تعتمد على محتوى

لم يغطيه الكتاب المقرر الذي يدرسه، وأن الصيغة الأخرى أنسب للمنهج الذي يدرسه.

جـ) شعر السيد باركر أن الاختبار أُعطي في وقت مبكر، واقترح تأخير الاختبار لعدة أيام، إذ سيكون عندها أداء عدة مفحوصين مختلفاً بشكل أساسي.

د) يرغب مستر وايز جمع درجات اختباري الرياضيات في مركب واحد لأنه يعتقد أنهما يقيسان الشيء نفسه.

هـ) يريد مرشد تربوي معرفة ما إذا كان تباين الخطأ في درجات المفحوصين أكبر أما بسبب الصيغة الاختبارية أو وقت الاختبار (والطلبة في مدرسته لم يتم اختبارهم بالصيغة نفسها، وبعضهم تم اختباره بأسبوع قبل غيرهم).

و) دوّن دليل اختبار في الشخصية ثبات التجزئة النصفية محسوباً بالصيغة:

$$\frac{4\,P_{GH}\;\hat{\sigma}_G\;\hat{\sigma}_H}{\hat{\sigma}_{X2}}\;\rho_{XX}\backslash$$

حيث تمثل G و H درجات نصفي الاختبار أثبت أن هذه الصيغة تكافيء معامل رولون.

6/ ابتداءً بصيغة KR20 المبينة في هذا الفصل، اشتق KR21 مع كتابة أي افتراض يلزمك في عملية الاشتقاق.

7/ قارن بين طرائق التجزئة النصفية والتباين المشترك للفقرات المختلفة المستخدمة في حساب الثبات بدلالة البيانات المطلوبة، وفسر لماذا نحصل على نتائج مختلفة ببعض هذه الطرائق.

الفصل الثامن

8

مدخل الى نظرية التعميم

الفصل الثامن

مدخل الى نظرية التعميم

تعد الطرائق المقدمة في الفصل السابع تقنيات فعالة في دراسة الثبات إلا أنها ليست بالمرونة الكافية لتتناسب مع مشكلات الثبات جميعها التي تظهر في الاختبارات العقلية. افترض مثلاً أن (50) طالباً كتبوا مقالات صححت على تدريج مؤلف من (10) نقاط بوساطة ثلاثة مصححين، ولأن التقنيات المبينة في الفصلين (7,6) تستخدم في حالة توافر قياسين لكل مفحوص، لذلك فإنه لا يمكن تطبيقها مباشرة قي هذا الموقف. ومثال آخر، استخدم محلل نفسي يعمل مع الأطفال من ذوي الاحتياجات الخاصة صيغتين متبادلتين لاختبار تشخيصي، واهتمامه منصب إلى الحد الذي تعطي فيه نتيجة الصيغتين الدرجة نفسها لكل طفل فضلاً عن انحراف الدرجات عن وسطها حتى يصل إلى صيغتين متوازيتين تماماً. ومن خلال النظرية التقليدية فإن كلاً من معامل الثبات والخطأ المعياري قد تعطي نتائج مضللة عندما تطبق في هذا الموقف، وعندما يتم حسابهما لاختبارين ليسا متوازيين تماماً فإن معامل الثبات والخطأ المعياري للقياس يقيسان اتساق الدرجات الزائية وانحراف الدرجات. ومثال أخير، تم توزيع اتجاهات المسترشدين في برنامج إعادة تأهيل المدمنين إلى مجموعتين ارشاديتين مختلفتين من خلال أربعة مرشدين. ويفيد هنا معرفة مساهمة كل من التباين عبر المرشدين والتباين بين المجموعتين في التباين الخطأ، ثانية فإن نظرية القياس التقليدية تزودنا بعنصر خطأ غير مميز لا يمكن تطبيقه بسهولة في هذه المشكلة.

أحد الأساليب المثمرة في حساب معاملات الثبات وتباين الخطأ في مثل هذه المواقف تعتمد على تحليل التباين. فاستخدام تحليل التباين في حساب معاملات الثبات وأخطاء القياس له تاريخ يعود إلى الأربعينات من القرن السابق. فمعظم الصيغ الأساسية وضعت من قبل بعض المؤلفين أو أنها واضحة من خلال أعمالهم، مثل هويت (Hoyt, 1951) وليندكويست (Lindquist, 1953) وليندي وميتزل (Medley& Mitzel, 1963)، وقد نشر كل من كرونباخ وجليسر وراجاراتنام (Cronbach, Gleser & Rajaratnam, 1963) أول بحث لهم في نظرية إمكانية التعميم. وكانت المساهمة الرئيسة لهذه الدراسات ليست تطوير صيغ جديدة للثبات وإنما تطوير طريقة للتفكير في الثبات. و/أو تباين الخطأ الأكثر ملائمة للموقف الذي سنتناوله.

ويقدم هذا الفصل مفاهيم نظرية إمكانية التعميم المعروفة بالتصميمات ذات البعد الواحد

والتصميمات ذات البعدين، وستجد معالجة أكثر شمولية لهذه النظرية في كرونباخ وزملاؤه (1972). ومصدر آخر أكثر شمولية من هذا الفصل ولكنها أقل شمولية وتعقيداً من كرونباخ وزملاؤه تجدها في بريتان (Brenan, 1983).

دراسات G ودراسات D:

من الضروري عند بدء دراسة نظرية إمكانية التعميم التمييز بين دراسات إمكانية التعميم (G) ودراسات القرار (D). فالباحث الذي يجري دراسة G يهتم بصورة أولية بمدى التعميم من عينة القياس إلى نطاقه. ويحدد النطاق عادةً من خلال مجموعة شروط القياس الأكثر شمولاً من الشروط التي تم فيها القياس. ويمكن اعتبار الدراسات التي تهتم باستقرار الاستجابات عبر الزمن، وتكافؤ درجات صيغتين أو أكثر للأداة نفسها أو العلاقات الداخلية بين درجات المقاييس الفرعية دراسات G ومن جهة أخرى فإن دراسات D هي التي تستخدم البيانات لهدف معين من أجل اتخاذ قرار، فهي تزود ببيانات تصف المفحوصين من أجل الاختيار أو التسكين أو مقارنة المجموعات في تجربة أو اختبار علاقة بين متغيرين أو أكثر.

وتستهدف دراسة G في المساعدة في التخطيط لدراسة D التي لها إمكانية تعميم ملائمة. لذا فإن تصميم دراسة G تحتاج المشاركة التامة في التصميمات التي يمكن استخدامها لدراسة D. ولا يمكن تصنيف الدراسة على أنها دراسة G أو D اعتماداً على تصميمها فقط، فهدف الفاحص هو العامل المحدد: افترض أن باحث أجرى اختباراً على 2000 طفل اختيروا عشوائياً من المدارس العامة و 2000 طفل آخرين اختيروا عشوائياً من المدارس الخاصة في المنطقة نفسها باستخدام اختبار تحصيلي مقنن واحد. فإذا كان هدف الباحث تحديد ما إذا كان الاختبار له الثبات نفسه، لكلا المجموعتين فيمكن تصنيف الدراسة على أنها G، وأما إن كان هدفه مقارنة متوسط مستويات تحصيل المجموعتين لعمل استنتاجات حول ملائمة النظام التربوي لكلا النظامين تصنف الدراسة على أنها D. كمبدأ على الفاحص أن يؤسس طريقة قياس تستخلص منها بيانات قابلة للتعميم قبل الشروع في دراسة D. وعملياً يمكن اختبار امكانية تعميم البيانات أحياناً باستخدام بيانات دراسة D، ولكن يجب إلغاء الاختصارات في السلوكات البحثية كلما أمكن ذلك.

وفي نظرية إمكانية التعميم تسمى شروط القياس أبعاد (Facets) فمثلاً، افترض أنه تم تدريج أداء عمال مصنع من خلال مشرفين أثنين وضمن عبء عمل: ثقيل، ووسط، وخفيف. هنا شروط القياس (الأبعاد) التي أُخذت فيها القياسات هي الإشراف وعبء العمل. توضيح آخر: أجرى باحث قياس للكتابة الإنشائية للأطفال في أربعة مناسبات، كتب كل طفل الإنشاء

في موضوعين مختلفين، في كل مناسبة، وتم تصحيح المواضيع من قبل ثلاثة مصححين. هنا استخدم التصميم ثلاثة أبعاد هي: المناسبات والموضوع والمصحح.

ويمكن معالجة البعد في دراسة D على أنه ثابت أو عشوائي، فعندما يهدف الباحث التعميم إلى الظروف التي تظهر فيها دراسة D فقط فإن البعد يكون ثابتاً. أما عندما يأخذ الباحث شروط دراسة D على أنها عينة من عدد كبير من الشروط ويهدف التعميم إلى الشروط الأخرى كافة فإن البعد يكون عشوائياً. وفي كلا الحالتين يطلق على القياسات التي تم الحصول عليها تحت شروط القياس جميعها اسم نطاق التعميم (Universe of generalization) وفي دراسة G أيضاً يجري القياس ضمن مجموعة لشروط محددة، والتي تؤخذ بعين الاعتبار لتمثل مجموعة أكبر من الشروط. ويسمى مجتمع الملاحظات التي تتضمن شروط القياس جميعها بـ «نطاق الملاحظات المسموح بها». وقد تكون نتائج دراسة G مفيدة للباحث الذي يخطط لدراسة D فقط في حالة كون نطاق الملاحظات المسموح يتضمن نطاق إمكانية التعميم.

تعرف الدرجة الحقيقية في نظرية القياس التقليدية على أنها متوسط درجة عدد كبير من الاختبارات المتوازية تماماً؛ وأما تباين الدرجة الحقيقية فهو تباين هذه المتوسطات جميعها. ويعرف الثبات على أنه نسبة تباين الدرجات الحقيقية إلى تباين الدرجات الملاحظة. وأما في نظرية إمكانية التعميم فإن درجة المفحوص الشاملة تعرف بأنها متوسط القياسات في نطاق التعميم، ولا يفترض أن تكون هذه القياسات متوازية تماماً. وإحدى طرق تعريف معامل إمكانية التعميم على أنه نسبة تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة المتوقع، وهذا الأخير يعتمد على تصميم دراسة D، لذا فإنه يوجد معاملات إمكانية تعميم مختلفة تناسب تصميمات دراسات D المختلفة (تعريف آخر لمعامل إمكانية التعميم: هو أنه مربع معامل الارتباط بين درجات النطاق والدرجات الملاحظة، وسيؤجل مناقشة هذا النوع من معاملات التعميم إلى آخر الفصل). أيضاً تعتمد درجة النطاق وبالتالي تباينها على نطاق التعميم. لذلك عندما يريد باحثين أثنين استخدام تصميم دراسة D نفسه ولكن لنطاقات تعميم مختلفة فهذا يتطلب معاملات تعميم مختلفة.

وفي الجزء التالي سيتم وصف كيفية صياغة معاملات إمكانية التعميم وحسابها لأربعة تصميمات جميعها ذات بعد واحد.. ومع أن الكثير من تطبيقات نظرية إمكانية التعميم تستخدم أبعاد عدة، ومن الأسهل فهم نظرية إمكانية التعميم إذا ركزنا أولاً على دراسات تستخدم بعداً واحداً فقط. وفي العرض التالي سنقدم مثال يستخدم تدريج سلوك لتوضيح صياغة معاملات إمكانية التعميم ومن ثم حسابها، وفي هذا الموقف يكون «المقدرين» هو البُعد الواحد. ومع ذلك يمكن تطبيق الطرائق الموصوفة في الأجزاء التالية على مواقف يتضمن بُعد القياس فيها صيغة الاختبار والمواقف الاختبارية ومطبقي الاختبارات.

صياغة معاملات إمكانية التعميم للتصميمات أحادية البُعد

افترض أن عالم نفس اجتماعي مهتم في دراسة ارتباطات الحذر، وأجرى دراسة تم خلالها ملاحظة تلميذين أثنين في تفاعلهم في موقف اجتماعي. وهذه التجربة عبارة عن دراسة D كان أحد التلاميذ هو مركز الملاحظة والآخر كان يعمل مع عالم النفس دون علم التلميذ الأول بذلك. هذا وسيتم تدريج مستوى الحذر على متصل مؤلف من (10) نقاط. وخطط عالم النفس لأن يكون التلميذ مساعداً في تدريج متصل قياس مستوى الحذر، وحتى في الموقف البسيط هذا هنالك عدة تصميمات محتملة لدراسة D. أربعة منها تكون على النحو الآتي:

1/ أعطي كل مفحوص درجة من قبل مقدر واحد، هو نفسه للمفحوصين جميعهم.

2/ أعطي كل مفحوص درجة من قبل مجموعة مقدرين، هم نفسهم للمفحوصين جميعاً.

3/ أعطي كل مفحوص درجة من قبل مقدّر مختلف، ويوجد مقدّر واحد لكل مفحوص.

4/ أعطي كل مفحوص درجة من قبل مجموعة مقدرين، ويوجد مقدرين مختلفين لكل مفحوص.

وفي التصميمات الأربعة جميعها يوجد بُعد واحد للمقدرين، ففي التصميمات الأول والثاني تعرض كل مفحوص للمقدرين نفسه، أي أنهم تعرضوا إلى شروط قياس نفسها، ويقال في مثل هذه الحالة أن البُعد تقاطع (Crossed) مع المفحوصين. ويرمز لهذا التصميم بـ "Pxi" وترمز P إلى الأفراد أو المفحوصين، وترمز i إلى شروط القياس، ويطلق على البيانات الناجمة عن مثل هذا التصميم اسم بيانات مزاوجة. ويتعرض كل مفحوص في التصميمين الأخيرين لمقدرين مختلفين أو بصيغة أكثر عمومية تعرضوا لشروط قياس مختلفة، ويقال هنا أن شروط القياس تداخلت "nested" ضمن المفحوصين. ويرمز لهذا التصميم بـ: "i:p"، وتقرأ i تداخلت ضمن P، وفي مثالنا i تداخل المقدرين ضمن المفحوصين، ويطلق على البيانات الناجمة عن هذا التصميم اسم بيانات مستقلة. وتعتمد الضيغة الملائمة لمعامل إمكانية التعميم على التصميم الذي يختاره الباحث عند تخطيطه لدراسة D. وكل معامل إمكانية تعميم عبارة عن نسبة تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة، لذلك تختلف قيم معاملات التعميم المرتبطة بكل تصميم لاختلاف تباين الدرجات الملاحظة المرتبط بكل تصميم. ويعتمد حساب معاملات إمكانية التعميم الأربعة على تصميم دراسة G. وفي البداية سنتناول حساب العامل المرتبط بتصميم دراسة G (2)، وسنستخدم في حساب هذه المعاملات تحليل التباين ثنائي العامل وملاحظة واحدة في كل خلية، ويطلق على هذا النوع من

تحليل التباين أيضاً اسم تحليل التباين المتكرر. ويقدم شافيلسون (Shavelson, 1981) في الفصلين 16 , 17 شرحاً واضحاً لتحليل التباين، كما يقدم مايرز (Myers, 1979) في الفصل السادس توضيحاً مناسباً لتحليل التباين المتكرر.

وقبل أن نعود لمناقشة نظرية إمكانية التعميم يجب ملاحظة أن تصميم 1، يمكن اعتباره حالة خاصة للتصميم 2، وذلك لأنه يستخدم مقدراً واحداً بدلاً من عدة مقدرين. كذلك فإن التصميم 3 ومعامل إمكانية تعميمه حالات خاصة للتصميم 4 ومعامل إمكانية تعميمه، ومع ذلك سنحتفظ بعرض أربعة تصميمات متميزة أثناء العرض.

معامل إمكانية التعميم للتصميم 1:

تذكر مثال عالم النفس الاجتماعي الذي يجري بحثاً يتطلب جمع تقديرات عن الحذر (تقدير الحذر). وافترض أن هذه الدراسة ستجري في الفصل الدراسي القادم، وأن عالم النفس غير متأكد أي التلاميذ سيساعده في عملية التقدير، وبصرف النظر فإنه مهتم باستخدام بيانات دراسة G لحساب إمكانية تعميم التقديرات التي يحصل عليها في دراسات D مستقبلاً. من وجهة نظر نظرية إمكانية التعميم فإن موقف عالم النفس كما يأتي: للاعتبارات التطبيقية سيجري مقدّر واحد تدريج المفحوصين جميعاً في دراسة D. وإن لم يكن مهماً من الناحية التطبيقية فإنه سيرتب المفحوصين بحيث يتم تقديرهم بوساطة عدد كبير جداً من المقدِّرين المدربين، وهذا التجمع من المقدرين يعد نطاق إمكانية التعميم. وبما أن الباحث استخدم مقدراً واحداً فإنه يود معرفة مدى جودة تعميم التقدير الذي أجراه مقدر واحد بالنسبة إلى متوسط التقديرات جميعها التي يؤشرها المقدرين جميعاً في نطاق إمكانية التعميم.

وقبل عرض طريقة حساب معامل إمكانية تعميم المقدرين سنحدد عدة كميات. تمثل Xpi درجة المفحوص p من قبل المقدر i، وكما لوحظ في الفصل السادس فإن الدرجة الملاحظة يمكن تصورها واحدة من عدة تقديرات ممكنة التي حددها المقدّر i للمفحوص p، لذلك فإن Xpi تعد متغير عشوائي. وتمثل Tpi المتوسط أو القيمة المتوقعة لهذا المتغير العشوائي، وهو الدرجة الحقيقية للمفحوص p عندما تم تقديره من قبل المقدر i. ونموذج التقدير الملاحظ ببساطة هو:

$$E_{pi} + T_{pi} = X_{pi} \quad \text{..........} \quad (1\text{-}8)$$

وهو نفسه في نظرية القياس التقليدية. والسؤال حول كيفية تعميم X_{pi} إلى T_{pi} هو

السـؤال الذي يطرح على ثبـات X_{pi}، وهذا ليس السـؤال الذي يهم عـالم النفس في الوقت الحالي، وللتلميذ p تتغير T_{pi} عبر المقدرين، ومتوسط T_{pi} عبر المقدرين هو الدرجة الشـاملة للمفحوص p، وهذا المتوسط يرمز له بالرمز μp. وسؤال إمكانية التعميم لعالم النفس يتعلق بكم الجودة في التقدير في دراسة D الذي سيعمم إلى μp: الدرجات الشاملة. كذلك نحتاج إلى تحديد μ (القيمة المتوقعة للدرجات الحقيقية للمفحوصين الذين تم تقديرهم من قبل المقدِّر i، أي أن μ هي المتوسط عبر المفحوصين لـ T_{pi}. وفي النظرية التقليدية تتساوى القيمة المتوقعة للدرجات الملاحظة مع القيمة المتوقعة للدرجات الحقيقية. كذلك في نظرية إمكانية التعميم تتساوى القيمة المتوقعة لـ X_{pi} (تقديرات المقدِّر i) والقيمة المتوقعة لـ T_{pi} (الدرجات الحقيقية للمفحوصين المقدّرة بوساطة المقدّر i). كذلك نحتاج إلى تعريف μ التي ترمز للمتوسط عبر التلاميذ في درجات النطاق، وهذا المتوسط الأخير يساوي المتوسط عبر المقدرين لـ μi، والتي يشار إليها بالمتوسط الإجمالي أو الكلي. ومن خلال هذه الرموز فإن النموذج الخطي لـ X_{pi} هو:

$$X_{pi} = \mu + (\mu p - \mu) + (\mu i - \mu) + e_{pi} \quad \ldots.. \quad 2\text{-}8$$

وبإعادة ترتيب المعادلة (2-8) للحصول على انحراف كل مفحوص عن المتوسط الاجمالي نصل إلى:

$$X_{pi} - \mu = (\mu p - \mu) + (\mu i - \mu) + e_{pi}$$

نلاحظ أن انحراف درجة المفحوص عن المتوسط الاجمالي مؤلف من ثلاثة عناصر: أثر المفحوص $\mu p - \mu$ وأثر المقدّر $\mu i - \mu$، والباقي e_{pi}. ويجب أن يلحظ القارئ اختلاف e_{pi} عن خطأ القياس E_{pi} في النظرية التقليدية، إذ أن e_{pi} يساوي E_{pi} إضافة إلى عنصر يعكس حقيقة أن الدرجات الحقيقية للمقدرين المختلفين ارتباطها غير تام. وأخيراً يلزمنا تعريف عناصر تباين عديدة:

σp^2_i : تباين درجات المفحوصين الشاملة، μp.

σ^2_i : تباين متوسطات المقدرين.

$\sigma^2_{e/i}$: تباين e_{pi} للمقدر i.

σ_e^2 : المتوسط عبر جميع المقدرين $\sigma^2_{e/i}$.

$\sigma^2_{x/i}$: تباين X_{pi} للمقدر i.

والآن بعد تحديد الرموز الضرورية دعنا نعيد النظر في دراسة G وكيفية استخدام نتائجها في تقدير معامل إمكانية التعميم الذي يناسب دراسة D التي سيجريها عالم النفس في الفصل الدراسي التالي.

افترض أن عالم النفس أجرى دراسة G على عشرة مفحوصين وتم تقدير الحذر لكل منهم من قبل ثلاثة مقدرين، والبيانات موضحة في جدول (8-1). مع ملاحظة أن دراسة D ستجري بوساطة مقدر جديد، فكيف يمكن استخدام هذه البيانات في حساب معامل إمكانية تعميم هذه التقديرات على دراسة D؟

جدول 8-1: درجات المفحوصين العشرة بوساطة ثلاثة مقدرين

رقم المفحوص	المقدرين			المتوسط (X_{p2})
	1	2	3	
1	2	3	2	2.33
2	8	5	7	6.66
3	4	2	2	2.66
4	4	3	6	4.33
5	8	5	5	6.00
6	8	5	7	6.66
7	6	4	5	5.00
8	4	3	3	3.33
9	3	2	2	2.33
10	1	2	3	2.00
المتوسط (X_{pi})	4.8	3.4	4.2	(X_{pi}) 4.13

تذكر أن الثبات يحدد في النظرية التقليدية على أنه نسبة تباين الدرجة الحقيقية إلى تباين الدرجة الملاحظة. ويبدو منطقياً الآن تحديد معامل إمكانية تعميم التقديرات في دراسة D على أنه:

$\sigma p^2/\sigma^2_{X/i}$: نسبة تباين الدرجة الشاملة إلى تباين الدرجة الملاحظة للمقدر المستخدم في دراسة D. وهنالك مشكلة تطبيقية مع هذا الاقتراح، وهو أن دراسة D ستطبق في الفصل التالي، وعالم النفس لا يدري أي المقدرين سيستخدم في دراسة D، إضافة إلى أن هذا المقدر قد لا يكون أحد المقدرين المستخدمين في دراسة G. لذلك لا يمكننا باستخدام بيانات دراسة G حساب $\sigma^2_{X/i}$ للمقدر الذي سيستخدم في دراسة D.

مع ذلك، إذا افترضنا أن المقدّر في دراسة D ممثل بوساطة مجموعة المقدرين في نطاق التعميم، فإن أفضل تخمين لتباين متوسط (أو القيمة المتوقعة) الدرجة الملاحظة ($\sigma^2_{X/i}$) للمقدرين الثلاثة في النطاق، ويمكن تبيان أن متوسط الدرجة الملاحظة يساوي $\sigma e2 + \sigma p2$. إضافة إلى أن بيانات المقدرين الثلاثة متوافرة من درسة G فإنه يمكن استخدامها لحساب $\sigma_e^2 + \sigma_p{}^2$. ونحصل على معامل إمكانية التعميم بتعويض تباين الدرجة الملاحظة للمقدر i بتباين الدرجة الملاحظة المتوقع للمقدرين جميعاً في نطاق التعميم، وكما يأتي:

$$\sigma^2_{i*} = \frac{\sigma^2 p}{\sigma^2_i + \sigma p^2} \qquad (3 - 8) \ldots\ldots\ldots\ldots$$

تستخدم الإشارة * في p^2_{i*} لتشير إلى أن المعامل مناسباً لدراسة D ضمن شروط التقاطع مع المفحوصين. لاحظ الافتراض الحرج في استخدام هذا المعامل كما هو الحال للمعاملات الأخرى المقدمة في هذا الفصل، وهو أن المقدرين في دراسة G ودراسة D يمثلون نطاق التعميم نفسه. ويكون هذا ممكناً فيما لو تم تحديد المقدرين في كلا من النطاق والعينة عشوائياً. وفي مثالنا هذا ليس ممكناً، لذلك فإن قضية كون المقدرين عينة ممثلة تصبح مسألة حكمية.

وحالما يتم تحديد معامل إمكانية تعميم من خلال عناصر التباين يجب اقتراح طريقة عملية لتقدير المعامل، وهذا ممكناً باستخدام تحليل التباين ثنائي العامل لبيانات جدول (8-1) والتي يمكن تصنيفها من خلال بُعدين هما: المفحوصين والمقدرين (وكل منهما يعبر عنه بعامل في تحليل التباين)، وكل مقدر يؤلف أحد مستويات بعد المقدّر، وكل تجمع لمفحوص ومقدِّر يسمى خلية. ويوصف التصميم المبين في جدول (8-2) على أنه تصميم ثنائي العامل بملاحظة واحدة في كل خلية لأنه يوجد درجة واحدة فقط لكل تجمع مفحوص–مقدر.

ويمكن تلخيص نتائج تحليل التباين ثنائي العامل بملاحظة واحدة في كل خلية في جدول يماثل جدول (8-2)، والاختصار S V في هذا الجدول يبين مصدر التباين، والاختصار SS يرمز لمجموع المربعات، والرمز df يمثل درجات الحرية، و MS يرمز إلى متوسط المربعات. ويتبين من الجدول إن MS تحسب بقسمة SS على df المناظرة لها، ويشير كل من np و ni إلى عدد المفحوصين وعدد المقدرين على التوالي في دراسة G. والجدول (8-3) يبين خلاصة حسابية لبيانات جدول (8-1).

جدول (8-2): الصيغ الحسابية ومتوسط المربعات المتوقع في تحليل التباين ثنائي العامل:

EMS	MS	df	SS	Sv
$\sigma_e^2 + n_i \sigma^2_p$	SSp/(np-1)	np-1	$n_i \sum_p (x_{pI} - x_{pi})^2$	المفحوصين (p)
$\sigma_e^2 + n_p \sigma^2_i$	SSi/(n-1)	ni-1	$np \sum_i (x_{p2} - xp_L)^2$	القدرين (i)
σ_e^2	$\frac{SSr}{(np-1)(n_i-1)}$	$(np-1)(n_i-1)$	$\sum_i \sum_i (x_{pi} - xp_L)^2$ -SSp-SSi	البواقي

$X_{pI} = \sum_i X\, p_i/n_i$ $\quad X_{pi} = \sum X_{pi}/np$ $\quad X_{pI} = \sum_i \sum_p X_{pi/ninp}$

إحدى طرق حساب p^2_ι تتطلب حساب عناصر التباين $\sigma^2 p$ و $\sigma^2 e$ باستخدام مجاميع موزونة للقيم في عمود MS في جدول خلاصة تحليل التباين. وبعدها يتم تعويض هذه القيم بالمعادلة (8-3). ولتحديد صيغة حساب عنصر التباين على الفاحص أن يختبر صيغ متوسط المربعات المتوقع (EMS) لكل مصدر تباين في جدول تحليل التباين. وهذه الصيغ تحدِّد العناصر النظرية للتباين التي تصاحب المربع المحسوب عددياً في جدول (8-3). وتحديداً فإن كل EMS تؤشر الصيغة للمجتمع التي تقدر من خلال قيمة مربع المتوسط المناظر في جدول الخلاصة، لذا فإن قيمة مربع المتوسط المتوقع لعامل الأفراد EMSp يحسب من مربع متوسط الأفراد MSp. والآن افترض أنك تريد حساب $\sigma^2 p$ من النتائج المبينة في جدول (8-3)، وبالحدس يمكن رؤية أن MSp تقدر قيمتها من $\sigma^2 p$ فقط. ومع ذلك فإن الصيغة في عمود EMS تبين أن قيمة MSp هي تقدير لـ $n_i\sigma^2 p + \sigma^2 e$، وكذلك فإن قيمة MSp لا يمكن استخدامها كتقدير لـ$\sigma^2 p$. وبالتمحيص المتعمق لصيغ EMS الأخرى في الجدول تكشف أن MSr هي تقدير لـ $\sigma^2 e$. وحيث أن:

$\sigma^2 e + n_i\sigma^2 p$ ، ويمكن تعويض $\sigma^2 e - n_i\sigma^2 p + \sigma^2 e = n_i\sigma^2 p$

بـ EMSp و $\sigma^2 e$ بـ EMSr،

جدول (8-3): خلاصة جدول تحليل التباين

SV	SS	df	MS	EMS
المفحوصين (p)	92.794	9	10.310	$n_i\sigma^2_P + \sigma^2_e$
المقدرين (I)	9.866	2	4.933	$n_p\sigma^2_1 + \sigma^2_e$
البواقي (R)	18.780	18	1.043	σ^2_e

ونتوصل إلى أن:

$$EMS_{r} - EMS_p = n_i\sigma^2 P$$

وكذلك

$$\frac{EMS_{r} - EMS_p}{n_i} = \sigma^2 P$$

وهذه الصيغة تبين أن عنصر التباين في صيغة EMS يمكن عزلها وذلك بإيجاد مجاميع خطية موزونة لصيغتين أو أكثر من صيغ EMS في جدول تحليل التباين. وحالما يتم تحديد التجمع المناسب فإن الحسابات للعينة في عمود MS تعوض في صيغة المجتمع لتقدير عناصر التباين وبتعويض عناصر EMS بعناصر MS نحصل على:

$$\frac{MS_{r} - MSP}{n_i} = \hat{\sigma}^2 P \qquad (4 - 8) \;..........$$

وبالتعويض العددي في جدول (8-3) في المعادلة (8-4) نحصل على:

$$3.089 = \frac{1.043 - 10.310}{3} = \hat{\sigma}^2 P$$

وصيغة σ^2_e ببساطة تساوي EMSr ولذلك فإن:

$$MSr = \hat{\sigma}^2 P \qquad (5-8) \;.............$$

وفي مثالنا $\sigma^2 P = 1.043$، ومن حساب كل من $\hat{\sigma}^2_p$ و $\hat{\sigma}^2_e$ يمكننا حساب P^2_i باستخدام الصيغة:

$$\frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2} = \hat{P}^2_i * \qquad (6-8) \;...........$$

وبالتعويض في المعادلة (8-6) نحصل على:

$$0.75 = \frac{3.089}{1.043 + 3.089} = \hat{\rho}^2{}_i{}^*$$

ومعادلة بديلة للتعبير عن $\hat{P}_i{}^2$ من خلال عناصر MS المختلفة هي:

$$\frac{MS_r - MS_p}{(n_i - 1)\, MS_r + MS_p} = \hat{\rho}^2{}_i{}^* \quad \text{.......... (4 - 8)}$$

وبالتعويض في المعادلة (8-7) نحصل على

$$0.75 = \frac{1.043\text{-}10.310}{1.043\ x(1\text{-}3) + 10.310} = \hat{\rho}^2{}_i{}^*$$

هذا ويجب أن تتفق نتائج تطبيق كلا المعادلتين (8-6) ,(8-7) ضمن حدود الخطأ، والاختيار بين المعادلتين هو نوع من الملائمة الحسابية. ومعادلات مماثلة للمعادلة (8-7) التي تعبر عن معامل إمكانية التعميم على أنه دالة لمتوسط المربعات تستخدم أيضاً في التصميمات (4,3,2)، إلا أنها غير معروضة هنا لأنه يفضل تقدير معامل إمكانية التعميم من خلال حساب عناصر التباين عدا عن شيوعها في أدبيات ومراجع القياس.

وفي المثال السابق لا يعرف عالم النفس أي المقدرين قام بعملية التقدير في دراسة D. افترض الآن أنه عرف الآن المقدِّر الذي أجرى عملية التقدير ويريد معرفة مدى جودة تعميم التقدير للدرجات الشاملة، وفي حالة وجود هذا المقدر في دراسة G. ماذا يمكن أن نعمل لحساب معامل إمكانية التعميم للتقديرات الخاصة بهذا المقدِّر؟ إن هذا المعامل يمكن حسابه فيما لو توافر على الأقل ثلاثة مقدرين في دراسة G، وكان أحدهم في دراسة D. بإعطاء هذا المقدر رقم (1) فإن معادلة حساب هذا المعامل بوجود ثلاثة مقدرين في دراسة G هي:

$$\frac{2(\hat{\sigma}_1\hat{\sigma}_3\hat{\rho}_{13} + \sigma^2{}_1\sigma^2{}_2\,\hat{\rho}_{12})}{4\,\hat{\sigma}_1\hat{\sigma}_2\,\hat{\sigma}_3\ \ \hat{\rho}_{23}} = \hat{\rho}^2(X_{pi}, \mu\rho) \quad \text{.......... (8 - 8)}$$

لاحظ أن المعادلة (8-8) تؤدي إلى تقدير مربع معامل الارتباط $\rho^2(X_{pi}, \mu\rho)$ بين تدريج المقدر (1) والدرجات الشاملة. وقد اشتق لورد ونوفيك **(Lord & Novick, 1968, P 268)** المعادلة 8-8 والتي يمكن تطبيقها على أكثر من ثلاثة قياسات. وباستخدام جدول (8-2) نحصل على:

$$\rho^2(X_{pi}, \mu_\rho) = \frac{^2[(0.790)(2.043)(2.57)+(0.914)(1.264)(2.573)]}{(0.825)(2.043)(1.264)(6.620)4}$$

$$= 0.90$$

تذكر أن $\rho^2_{\iota*} = 0.75$، وبالتالي فإن معاملي إمكانية التعميم قدما انطباعين مختلفين لدراسة D، وهذه الفجوة بين المعاملين لأن $\rho^2_{\iota*}$ تشير إلى اختيار مقدِّر عشوائياً، في حين أن $(\rho^2 X p_\iota, \mu_\rho)$ تشير إلى مقدِّر محدد. وعند هذه النقطة يحتار القارىء في معرفة متى يكون معامل إمكانية التعميم المختار في الظروف العشوائية أقل من المعامل المحدد بشروط مخصصة، وأيهما يفضل على الآخر؟

أولاً: يمكن ملاحظة أن معامل الشروط العشوائية قد يكون أكبر أو أصغر من معامل الشروط المخصصة. ويبدو أن كل من كرونباخ وجليسر وراجاراتنام (1963) ولورد ونوفيك (1968) متفقين على أن معامل الشروط المخصصة أفضل من الآخر نظرياً، مع أن كرونباخ وزملاؤه اهتموا بدقة تقدير هذا المعامل والذي يتطلب عدداً كبيراً من المقدرين في دراسة G. وكنتيجة لهذا الاهتمام فإنهم فضلوا حساب معامل الشروط العشوائية، إلا أن لورد ونوفيك يميلان ألى تفضيل حساب معامل الشروط المخصصة. مع أن كلا المصدرين يقدمان صيغاً عامة لتقدير هذا المعامل.

معامل إمكانية التعميم للتصميم الثاني:

كان معامل إمكانية التعميم (.75) في الموقف الذي لم يعرف فيه الباحث أي المقدرين سيظهر في دراسة D. ومع أن معامل إمكانية تعميم هذا المعامل أساسي إلا أن عالم النفس قد يكون مهتماً في المعامل الذي ينتج فيما لو أشار مقدرين أثنين أو أكثر أن الدرجة هي المتوسط عبر المقدرين في دراسة D، وهنا يكون معامل إمكانية التعميم المناسب:

$$\rho^2_{\iota*} = \frac{\sigma^2 p}{\sigma^2 e/n_\iota\backslash + \sigma^2 p} \qquad \text{.......... (9 - 8)}$$

ترمز $n_\iota\backslash$ في المعادلة إلى عدد المقدرين الذين يشكل متوسط تقديراتهم X_{pI}. لاحظ أن n_i هي عدد المقدرين في دراسة G في حين أن $n_\iota\backslash$ هي عدد المقدرين في دراسة D. والفرق الرئيسي بين الصيغتين لكل من $\rho^2_{\tau*}$ و $\rho^2_{\tau*}$ هو أنه في الأخيرة تقسم σe^2 على عدد المقدرين في دراسة D، وتم هذا الإجراء لأن المتوسطين عبر مقدرين عدة يقلل تباين الخطأ.

وتقدير $P^2_{I^*}$ يتقرن باستخدام نتائج تحليل التباين ثنائي العامل، والتقدير يتم من خلال الصيغة:

$$\hat{\rho}^2_{\iota*} = \frac{\hat{\sigma}^2_{\rho}}{n_{\iota}\backslash / \hat{\sigma}_e^{2} + \hat{\sigma}^2_{\rho}} \quad (10 - 8) ....$$

وثانية، تحسب حدود التباين من المعادلات (8-4 ,8-5) ووجدنا سابقاً أن $\hat{\sigma}^2 p = 3.089$، و $\hat{\sigma}_e^2$ 1.043. افترض الآن أن عالم النفس استخدام مقدرين أثنين $n_2\backslash = 2$ في دراسة D. بالتعويض في المعادلة (8-10) نحصل على:

$$\rho_{I^*} = \frac{3.089}{2/1.043 + 3.089} = 0.85$$

نجد أن زيادة عدد المقدرين إلى 2 يزيد معامل إمكانية التعميم من (.75 إلى .85). كذلك فإن حساب $\hat{P}_{i*}$ لقيم عديدة للمقدرين $n_i\backslash$ يوفر معلومات مهمة في اختيار عدد مناسب من المقدرين لدراسة D. ويمكن حساب المعامل $\hat{P}_{i*}$ أيضاً باستخدام الصيغة:

$$\rho^2_{i*} = \frac{MS_r - MS_p}{(n_i\backslash / MSr\ (n_i\backslash - n_i) + MSp} \quad (11 - 8) ....$$

وبالتعويض في المعادلة 8-11 ينتج:

$$\rho^2_{\iota*} = \frac{1.043 - 10.310}{2/(1.043)\ (2-3) + 10.310} = 0.86$$

وعندما تكون $n_i\backslash = n_i$ فإنه يمكن تبسيط المعادلة (8-11) لتصبح:

$$\hat{\rho}^2_{i*} = \frac{MSr - MSp}{MSp} \quad (12 - 8) ....$$

وهذه الصيغة واسعة الإنتشار في الأبحاث المتعلقة بالموضوع. وعرضياً نلاحظ أن $\hat{p}^2_{i*}$ و $\hat{p}_{i*}$ تربط بينهما علاقة (نبوءة) سبيرمان- براون، وكما يأتي:

$$\rho^2_{i*} = \frac{n_{\iota}\backslash \rho^2_{\iota*}}{\rho^2_i\ (1 - n_i\backslash) + 1}$$

وكما في التصميم الأول، هنالك معامل آخر يكون مناسباً في حالة كون المقدرين في دراسة D كانوا من ضمن المقدرين في دراسة G، ويرمز لهذا المعامل بـ $\rho^2 (X_{pI} , \mu_p)$ الذي يمثل مربع الارتباط بين درجة D الملاحظة ودرجة النطاق. وهنا تكون الدرجة الملاحظة للمفحوص مساوية: لمتوسط الدرجات المعطاة للمفحوص من قبل المقدرين في دراسة D. وكما في التصميم الأول يلزم لحساب معامل إمكانية التعميم مقدرين أثنين في دراسة G إضافة إلى المقدرين الذين سيظهروا في دراسة D. وعند إضافة عدد مقدرين أثنين يمكننا استخدام المعادلة (8-8) لحساب $\rho^2 (X_{pI} , \mu_p)$. وباستخدام المعادلة (8-8) في هذا الموقف فإن متوسط درجات المفحوص في دراسة G المُعطاة من قبل المقدرين الذين ظهروا في دراسة D تحسب لكل مفحوص. وهنا يرمز $\hat{\sigma_1}$ إلى الإنحراف المعياري للدرجات. والرموز $\hat{\rho}_{12}$, $\hat{\rho}_{13}$ تشير إلى الارتباط بين هذه الدرجات والدرجات المُعطاة من قبل المقدرين الأثنين في دراسة G الذين لم يظهروا في دراسة D.

معامل إمكانية التعميم للتصميم الثالث:

افترض أن ملاحظة وإعطاء الدرجات لكل مفحوص في التجربة استغرق وقتاً طويلاً، ولكن عند توافر عدد كبير من المساعدين الذين يمكن استخدامهم كمقدرين، وهنا قد يستخدم عالم النفس مقدرين مختلفين لتقدير كل مفحوص. تذكر أن المقدرين في هذه الحالة تداخلوا مع المفحوصين. دعنا الآن نأخذ تباين المفحوصين عبر المقدرين بعين الاعتبار، وحيث أن كل مفحوص قدّر من قبل مقدر مختلف فإن الفروق عبر المفحوصين تتأثر بالفروق عبر المقدرين كما تتأثر بتباين فروق درجة النطاق والبواقي، ونتيجة هذا فإن تباين الدرجة الملاحظة في التصميم 3 يكون $\sigma e^2+\sigma^2{}_i+\sigma^2 p$. ويعكس الحد $\sigma^2{}_i$ الفروق عبر المقدرين. ومعامل إمكانية للتعميم للتصميم 3 يكون:

$$\rho^2{}_i = \frac{\sigma \rho^2}{\sigma e^2+ \sigma^2{}_i +\sigma^2 p} \quad \text{..........} \quad (13 - 8)$$

لإعطاء إشارة استخدام $p^2{}_\iota$ في حالة تداخل ι في P، فإن i في $p^2{}_\iota$ لم يؤشر لها بإشارة *.

ويستخدم حساب $p^2{}_\iota$ في تصميم 3 لدراسة G تقديرات ρ^2_p وσ^2_e وبالمعادلات (8-4) و (8-5) على التوالي، إضافة إلى ذلك يلزم تقدير $\sigma^2{}_i$ (وكما سنرى في تصميم 4 لدراسة G فإن تقدير $p^2{}_\iota$ يتطلب تقدير منفصل للكمية $(\sigma e^2+\sigma^2{}_i)$. ويمكن اشتقاق $\sigma^2{}_i$ باستخدام EMS في جدول 8-3 وعلى النحو الآتي:

$$\hat{\sigma}^2_i = \frac{EMSr_ EMSi}{np}$$

كذلك فإن σ^2_i تقدر بـ :

$$\hat{\sigma}^2_i = \frac{MSr - MSi}{np} \quad \text{.......... (14 - 8)}$$

وباستخدام قيم متوسط المربعات في جدول (8-3) نحصل على :

$$\hat{\sigma}^2_i = \frac{1.043 - 4.933}{10} = 0.389$$

ومعادلة تقدير معامل إمكانية التعميم هو:

$$\hat{\rho}^2_i = \frac{\sigma\, p^2}{\hat{\sigma}e^2 + \hat{\sigma}^2_i + \hat{\sigma}^2p} \quad \text{.......... (15 - 8)}$$

وهذا يؤدي في مثالنا إلى:

$$\hat{\rho}^2_i = \frac{3.089}{1.043 + 0.389 + 3.089} = 0.68$$

من مقارنة معامل إمكانية التعميم للتصميم (3) (.68) مع معامل أمكانية التعميم للتصميم (1) (0.75) نجد انخفاضا في التعميم. مع ذلك فإن الإنخفاض ليس كبيراً. من هنا نقترح أنه في حالة إمكانية تطبيق التصميم (3) يمكن اعتبارهُ تصميماً منطقياً لتجربة عالم النفس. وكما هو الحال مع معاملات إمكانية التعميم الأخرى توجد صيغة لتقدير $\hat{\rho}^2_i$ مباشرة من MS:

$$\hat{\rho}^2_i = \frac{MS_r_MS_p}{p + ni\, MS_i/_{np} + MSp_(n_i np_n_i_n_p)\, MSr/} \quad \text{.......... (16 - 8)}$$

ومن المهم تمييز أن $\hat{\rho}^2_i$ معامل مناسب لتصميم دراسة D المتداخلة، ويمكن حسابه باستخدام بيانات تصميم دراسة G المتقاطعة. وهذا يوضح أن دراسة G المصممة تصميماً جيداً تسمح بحساب معامل إمكانية التعميم الذي يرافقها لمدى واسع من تصميمات دراسة D. ومقارنة معاملات إمكانية التعميم لتصميمات عديدة تسمح للباحث مقارنة ومقابلة تصميمات دراسة D من خلال إمكانية تعميمها واختيار تصميم دراسة D الممكن تطبيقه ويتسم بإمكانية تعميم مناسبة.

معامل إمكانية التعميم للتصميم الرابع:

افترض أنه كان بالإمكان تقدير كل مفحوص من قبل مقدرين مختلفين في دراسة D، ولكن يريد عالم النفس مقدرين أثنين أو أكثر لكل مفحوص. معامل إمكانية التعميم المناسب لهذا التصميم يكون:

$$\hat{\rho}^2_{L*} = \frac{\sigma p^2}{(\sigma e^2 + \sigma^2_i)/n_i^{\backslash} + \sigma^2 p} \quad \text{(17 - 8) ....}$$

ثانية يشير تذييل L في ρ^2_i إلى استخدام متوسط التقديرات لكل مفحوص وبمقارنة المعادلة (8-17) مع المعادلة (8-13)، نرى أن الحد $\sigma e^2 + \sigma^2_i$ مقسوماً على $n_i^{\backslash}$ في المعادلة (8-17)، لذا فإن استخدام عدد مضاعف من المقدرين يقلل من أثر تباين المقدرين وتباين البواقي في تباين الدرجة الملاحظة ويزيد من معامل إمكانية التعميم. الصيغة لـ ρ^2_L هي:

$$\hat{\rho}^2_L = \frac{\hat{\sigma} p^2}{(\hat{\sigma}^2_i + \hat{\sigma} e^2)/n_i^{I} + \hat{\sigma}^2 p} \quad \text{(18 - 8) ....}$$

ولمقدرين أثنين، ينتج عن هذه المعادلة:

$$\hat{\rho}^2_L = \frac{3.089}{2/(1.043 + 0.389) + 3.089} = 0.81$$

جدول (8-4): معاملات إمكانية التعميم لتصميمات دراسة D الأربعة للبعد الواحد

التصميم	الوصف	عدد شروط القياس	تباين الدرجة الملاحظ	معامل امكانية التعميم
1	Pxi	1	$\sigma^2 p + \sigma e^2$	$\rho^2_L = \frac{\sigma p^2}{(\sigma p^2 + \sigma e^2)}$
2	Pxi	$n_i^{\backslash}$	$n_i\sqrt{\sigma e^2} + \sigma^2 p$	$\rho^{2*}_L = \frac{\sigma p^2}{n_i^1/\sigma e^2 + \sigma p^2}$
3	P: i	1	$\sigma e^2 + \sigma^2 i + \sigma^2 p$	$\rho^2_L = \frac{\sigma p^2}{\sigma e^2 + \sigma^2 i + \sigma p^2}$
4	P: i	$n_i^{\backslash}$	$n_i\sqrt{(\sigma e^2 + \sigma^2 i)} + \sigma^2 p$	$\rho^2_L = \frac{\sigma p^2}{n_t\sqrt{(\sigma e^2 + \sigma^2 i)} + \sigma p^2}$

ولمثالنا هذا، يوجد صيغة لتقدير $p^2{}_L$ مباشرة من متوسط المربعات:

$$\rho^2{}_L = \frac{MS_r - MS_p}{n_p n_i / MSe\,(n_i - np n_i - np\, n_i) + n_i MS_i / np n_i + MSp} \quad \text{.......... (19 - 8)}$$

ملخص معاملات إمكانية التعميم للتصميمات الأربعة:

قدمنا صياغة لمعاملات إمكانية التعميم الأربعة، واحدة لكل تصميم. ويبين الجدول (8-4) تباين الدرجة الملاحظة ومعاملات إمكانية التعميم للتصميمات الأربعة. ونعود لنؤكد على حقيقة أن معاملات امكانية تعميم مختلفة ترتبط بكل تصميم دراسة D. ويمكن حساب المعاملات الأربعة فيما لو جمعت بيانات دراسة G باستخدام التصميم الثاني. وكما سنرى في الجزء التالي فإنه لو جمعت بيانات دراسة D باستخدام التصميم الرابع فإننا يمكننا حساب معاملات إمكانية التعميم للتصميمين الثالث والرابع $(\rho^2{}_L , \rho^2{}_i)$ فقط.

دراسة G لبعد واحد متداخل (التصميم الرابع):

من المكن إجراء دراسة G بتصميم 4 إذ يمكن قياس كل مفحوص ضمن شروط مختلفة عن الآخرين. ففي مثال التقدير يظهر تصميم 4 فيما لو تم تقدير كل مفحوص من قبل مجموعة مقدرين مختلفين. ويمكن استخدام دراسة G التي تعتمد تصميم 4 فقط لحساب معاملات إمكانية التعميم للدراسات التي تجري باستخدام التصميمات 3 و 4. ونتيجة لذلك فإن دراسات D مثل هذه تميل لأن تكون ذات قيمة أقل من تلك التي تجري بالتصميم 2. علاوة على أن دراسة G التي تجري بالتصميم 4 تتطلب صيغاً جديدة لتقدير هذه المعاملات وعندما تجمع بيانات دراسة G باستخدام تصميم 4 فإن صيغ حساب $P^2{}_l$ و $P^2{}_i$ تعتمد على تحليل التباين الأحادي، والعامل الوحيد هو المفحوصين، وتعالج تقديرات المفحوصين على أنها قياسات متكررة على المفحوصين. ويبين جدول (8-5) الصيغ اللازمة لحساب تحليل

جدول (8-5): الصيغ الحسابية ومربعات المتوسطات المتوقعة لتحليل التباين الأحادي:

SV	SS	df	MS	EMS
المفحوصين	$n_i \sum_p (X_{pl} - X_{pl})^2$	1-np	σ(np-1)/ SSp	$n_i\, \sigma^2 p + \sigma^2{}_i + \sigma^2 e$
البواقي	$\sum_p \sum_i (X_{pl} - X_p i)^2$	$(1\text{-}n_i)\, np$	$[(1-n_i)np]$ / SSr	$\sigma^2{}_i + \sigma^2{}_e$

$n_i n_p / X_{pi} \sum_i \sum_p = X_{pI}$ $\quad\quad$ $n_i / X_{pi} \sum = X_{pI}$

التباين الأحادي، وخلاصة جدول تحليل التباين مبينة في جدول (8-6). والقيم العددية المذكورة تم الحصول عليها بتطبيق الصيغ المناسبة على بيانات جدول (8-2)، وعولجت هنا كما لو أنها جمعت بالتصميم 4. والصيغ اللازمة لحساب عناصر التباين هي:

$$\frac{MS_r - MS_p}{n_i} = \hat{\sigma}^2 p \quad \text{.......... (8 - 20 أ)}$$

$$\text{و} \quad MS_r = \sigma^2 e + \sigma^2 i \quad \text{.......... (8 - 21 ب)}$$

وفي مثالنا، ينتج عن هذه الصيغ:

$$2.956 = \frac{1.432 - 10.310}{3} = \hat{\sigma}^2 p$$

$$\text{و} \quad 1.432 = \sigma^2 e + \sigma^2 i$$

ويمكن حساب المعاملات P_i^2 و P_I^2 باستخدام المعادلتين (8-15) و (8-18) على التوالي:

جدول (8-6): خلاصة جدول تحليل التباين

MS	df	SS	SV
10.310	9	92.794	المفحوصين
1.432	20	28.646	البواقي

النطاقات والأبعاد الثابتة:

في الأمثلة السابقة جميعها أهتم الباحث في التعميم على نطاق يتضمن شروط قياس غير تلك المستخدمة في دراسة D. ففي هذه الأمثلة كان بُعد القياس بُعداً عشوائياً، وافترض أن المصطلحات المشتقة من حقيقة أن شروط القياس في دراسة D أنها عينة عشوائية من الشروط الأكثر شمولاً والتي تؤلف نطاق التعميم. وفي أمثلة أخرى يرغب الباحث التعميم على نطاق يتضمن فقط الشروط التي تظهر في دراسة D. وفي هذه المواقف يكون بُعد القياس ثابتاً. وهذا الجزء يهتم بنظرية إمكانية التعميم لنطاقات ذات أبعاد ثابتة، ونوقشت دراسات D التي تستخدم التصميم (1 و 2)، إذ يبدو أنها الأكثر شيوعاً، ومع ذلك يمكن اشتقاق نتائج مشابهة للتصميمات (3 و 4).

معامل إمكانية تعميم التصميم 1/ دراسة D ببُعد ثابت:

استرجع تطوير النموذج الخطي لـ X_{pi} لرؤية مضامين معالجة البعد على أنه ثابت أو عشوائي. وقد أشرنا إلى أنه يمكن تصوّر X_{pi} على أنها إحدى التقديرات التي أعطاها المقدر i للمفحوص P. لذلك فإن X_{pi} تكون متغير عشوائي للمفحوص P والمقدر i ويرمز للمتوسط أو القيمة المتوقعة لـ X_{pi} بالرمز T_{pi}، وعرفناها على أنها متوسط T_{pi} عبر المقدرين.

افترض الآن وجود فاحصين أثنين حصلوا على بيانات من تصميم 1/ دراسة D. تذكر أنه يستخدم مقدر واحد في هذا التصميم «المقدر i»، ويجري تقديراً للمفحوصين جميعاً. والفاحص الأول تصوّر المقدر i على أنه ممثل لنطاق أكبر من المقدرين، أي أنه عدَّ المقدرين بُعد عشوائي ويريد معرفة مدى جودة تعميم X_{pi} على μ_p: درجة النطاق للمفحوص p في نطاق مقدرين عدة. ومعامل إمكانية التعميم المناسب هو :

$$\frac{\sigma p^2}{\sigma e^{2}+\sigma p^2} = p^2_{\iota*}$$

والفاحص الثاني لم يكن مهتماً بالتعميم إلى نطاق مؤلف من مقدرين عدة، عدا عن كونه نطاق تعميمه مؤلفاً من مقدر واحد. وكنتيجة فإن الفاحص مهتم بمدى جودة تعميم X_{pi} على T_{pi} أي على درجة نطاق المفحوص المقدرة من قبل المقدر i، وهنا يكون تباين درجة النطاق هو تباين X_{pi} ويرمز له بـ $\sigma^2_{x/i}$. ونتيجة لذلك فإن معامل إمكانية التعميم الثاني يكون $\sigma^2_{T/i}/\sigma^2_{x/i}$: أي نسبة تباين الدرجة الحقيقية إلى تباين الدرجة الملاحظة للمقدر i، وهذا يكافيء معامل الثبات التقليدي للمقدر الواحد. ونقطة مفتاحية- دليلية- هو أن تعريفات درجة النطاق ومعامل إمكانية التعميم تكون نفسها للأبعاد الثابتة والعشوائية. وحتى عند استخدام التصميم نفسه لدراسة D فإن تغيير البعد من عشوائي إلى ثابت يغير حدود النطاق وبالتالي يغير درجة النطاق، وتباينها ومعامل إمكانية تعميمها.

كذلك يمكن تحديد المعامل المناسب لتصميم 1/ دراسة D ببعد ثابت، والسؤال المهم هو ما إذا كان بالإمكان تقديره من البيانات المجمعة في دراسة G. والإجابة أنه لا يمكن حسابه ولكن يمكن تقدير الحد الأدنى والذي يتم باستخدام بيانات دراسة G/ تصميم 2،والذي يظهر فيها المقدر بكل من دراسة G ودراسة D،ومع هذا المقدر المشترك يمكن حساب $\sigma^2_{x/i}$ بسهولة، وذلك بحساب تباين التقديرات $\sigma^2_{x/i}$التي حدَّدها المقدِّر المشترك. ومع ذلك لا نستطيع في العادة

حســـاب $\sigma^2_{T/i}$، ولكن الحـــد الأدنى لـ $\sigma^2_{T/i}\ /\sigma^2_{x/i}$ يســـاوي $1 - \frac{\sigma_e^2}{\sigma^2_{x/i}}$ (Lord & Novick, 1968). وحسب هذا المعامل بتعويض $\hat{\sigma}_e^2$ مكان σ_e^2 و $\hat{\sigma}^2_{x/i}$ مكان $\sigma^2_{x/i}$ والكمية $\hat{\sigma}_e^2 = MSr$ والتي تحسب من تحليل التباين الثنائي لبيانات دراسة G. وللمقدر 1 في جدول (1-8) فإن المعامل يكون= $1 - \frac{1.043}{6.622} = 0.84$

معامل إمكانية تعميم التصميم 2/ دراسة D ببعد ثابت

أشرنا إلى أنه عند استخدام تصميم 1 لإجراء دراسة D وببعد ثابت يكون معامل إمكانية التعميم المناسب مكافئاً لمعامل الثبات في النظرية التقليدية. وبصورة مشابهة فعند استخدام تصميم 2 لإجراء دراسة D وببعد قياس ثابت، فإن معامل إمكانية التعميم المناسب ثابتة يكافيء معامل الثبات في النظرية التقليدية. وفي التصميم 2 فإن الدرجة الملاحظة هي متوسط الدرجات التي حصل عليها المفحوص، لذا فإن المشكلة تكمن في حساب ثبات الدرجة المركبة. وكما لوحظ في الفصل السابع اعتبر معامل ثبات α الحد الأدنى لمعامل الثبات لمثل هذا المركب، مع العلم بأن المقدرين الذين ظهروا في دراسة D هم فقط المقدرين في دراسة G. ويمكن اعتبار معامل ثبات α المحسوب في دراسة G هو الحد الأدنى لمعامل إمكانية التعميم المناسب.

استخدام نظرية إمكانية التعميم مع بيانات غير التقديرات:

يجب أن يكون واضحاً من هذا العرض أنه يمكن تطبيق نظرية إمكانية التعميم بشكل مفيد في بيانات التقدير. ومع ذلك فإنه يمكن تطبيقها على بيانات من أنواع أخرى من الفقرات والاختبارات الأقل وضوحاً. في المثال الأول افترض أن لدينا ثلاثة صيغ بديلة لاختبار تحصيلي معياري، وبما أنه يمكن بناء صيغ أخرى للاختبار نفسه فمن المنطقي تصوّر هذه الصيغ على أنها عينة ممثلة للصيغ المحتملة الممكن بناؤها. والآن افترض أن دراسة في إحدى المقاطعات تستخدم صيغة واحدة من الثلاث في دراسة D. ومن المهم معرفة مدى جودة تعميم درجات هذه الصيغة على نطاق الدرجات لصيغ الاختبار الممكن بناؤها جميعاً. في هذه الحالة فإن معامل إمكانية التعميم الأنسب هو $P(X_{pi}, \mu_p)$. ولحساب هذا المعامل يجب أن تجري المقاطعة دراسة G باستخدام الصيغ الثلاث، واستخدام المعادلة (8-8) لحساب معامل إمكانية التعميم. وكمثال ثاني: افترض أن عالم نفسي خطط لدراسة العوامل المرتبطة بالشعور

بالذات، والدراسة المخطط لها هي دراسة D. وطوّر الباحث أداة مؤلفة من 20 فقرة صح/خطأ وخطط لتطبيقها على المفحوصين جميعهم في دراسة D، وتصوّر الفقرات هنا على أنها عينة عشوائية من نطاق الفقرات الممكن استخدامها، ولاحظ أن الباحث يمكنه تسجيل درجات الفقرة للمفحوصين جميعهم في دراسة D في جدول مشابه للجدول (8-1) فالصفوف في الجدول لازالت تشير إلى المفحوصين ولكن الأعمدة تشير إلى الفقرات، لذا فإنه يمكن تصور تصميم دراسة D على أنه تصميم 2 وبفقرات تمثل مقياس البعد، وحيث أن الدراسة ستجري بتصميم 2، فإن معامل إمكانية التعميم المناسب يكون $\rho^2_{\iota*}$، والسؤال الذي يظهر هو: ما نوع دراسة G التي يجب أن يجريها الباحث لحساب $\rho^2_{\iota*}$. وتذكر أنه لزم في مثال التقدير التصميم 2 لحساب $\rho^2_{\iota*}$، وبصورة مناسبة ففي المثال الحالي يلزمنا دراسة G بتصميم 2. لذا فإنه يمكن للباحث أن يطبق الأداة على مفحوصين عددهم np، ويستخرج منها استجابات جدول $n_p \times n_\iota$ مشابه لجدول (8-1)، وباستخدام المعادلات (8-4) و (8-5) و (8-10) لحساب $\rho^2_{\iota*}$، وتكون قيم كل من n_ι و $2 = n_{\iota}'$.

ويمكن تبيان أن $\rho^2_{\iota*}$ تكافئ KR-20 للفقرات ثنائية التصحيح ببعد واحد عشوائي، والحد الأدنى لمتوسط $\rho^2(X_{pI}, \mu_p)$ يساوي $\rho^2_{\iota*}$. لذا يمكن للباحث تصوّر أن KR20 كحد أدنى لارتباط متوسطات المربعات مع درجة نطاق اختبارات مبنية بالاختيار العشوائي لفقرات عددها 20 من نوع صح- خطأ من نطاق الفقرات.

أخطاء القياس المعيارية للقرارات المطلقة والنسبية:

تستخدم الأخطاء المعيارية بشكل أساسي في تكوين فترة ثقة للقياس. ففي النظرية التقليدية يكون لفترة الثقة الصيغة $x \pm$ SEM، إذ ترمز x إلى الدرجة الملاحظة وترمز SEM إلى الخطأ المعياري للقياس كما حددتها النظرية التقليدية، وتعد هذه الفترة فترة ثقة للدرجة الحقيقية، ويكون الوضع أكثر تعقيداً بعض الشيء في نظرية إمكانية التعميم، وهذه الفترة لها الصيغة S±SEM، وهنا ترمز S إلى الدرجة المستخدمة في تكوين الفترة. وتعتمد الدرجة والخطأ المعياري بالأساس على نوع القرار المتخذ باستخدام بيانات دراسة D، وثانوياً على تصميم الدراسة أو عدد شروط القياس المستخدمة.

ففي دراسة D أحادية البعد هنالك أربعة تعريفات لتباين الدرجة الملاحظة، واحد لكل تصميم، ويمكن تحديد خطأ التباين المرافق للتصميم على أنه ذلك الجزء من تباين الدرجة الملاحظ والذي ليس تباين درجة النطاق. فالتصميم وتباين الخطأ المرافق له على النحو الآتي:

تصميم 1 : σ_e^2

تصميم 2 : $n_i' \sigma_e^2$

تصميم 3 : $\sigma_e^2 + \sigma_i^2$

تصميم 4 : $n_i\sqrt{(\sigma_e^2 + \sigma_i^2)}$

ويحدد الخطأ المعياري للقياس على أنه الجذر التربيعي لتباين الخطأ. ولا يعتمد تقرير استخدام أي الأخطاء المعيارية بالضرورة على تصميم دراسة D، علاوة على أنه يعتمد بشكل أولي في الاستخدام لمن ستوضع الدرجات في دراسة D. ومن المهم هنا التمييز بين القرارات المقارنة والقرارات المطلقة، فالقرار المقارن الذي يتخذ بناءً على مقارنة درجات المفحوصين المختلفين أمثلة على هذه القرارات التي تتضمن استخدام درجات الاختبار لترتيب الأفراد على منحنى واختيار المفحوصين ذوو الترتيب الأعلى ليدخلوا في معهد تربوي. ويعتمد الخطأ المعياري لقرارات المقارنة على تباين الخطأ لتصميم دراسة D، فمثلاً إذا أجريت دراسة D/ بتصميم 3 فإن الخطأ المعياري يكون $\sqrt{\sigma_e^2 + \sigma_i^2}$,

ويعتمد القرار المطلق على درجة المفحوص فقط لا على مقارنتها بدرجات المفحوصين الآخرين. فمثلاً يستخدم الاختبار محكي المرجع لتحديد ما إذا كان يمكن للمفحوص التقدم إلى الوحدة التالية في البرنامج التعليمي. واتخاذ القرار فيما يتعلق بتقدم المفحوص يحدث فيما لو تجاوزت درجة المفحوص محك محدد مسبقاً. فعندما تستخدم درجات الاختبار لاتخاذ قرار مطلق وباستخدام تصميم أحادي البُعد، فإن الخطأ المعياري للقياس يعتمد على عدد الشروط التي قيس المفحوص ضمنها، وعندما يكون هنالك شرطاً واحداً (تصميم 1 و 3) فإن تباين الخطأ يساوي $\sigma_e^2 + \sigma_i^2$، بينما لو كان هنالك عدداً من الشروط (تصميم 2 و 4) فإن تباين الخطأ يساوي $n_i\sqrt{(\sigma_e^2 + \sigma_i^2)}$. وكمثال على قرار مطلق افترض أنه تم استخدام اختبار مؤلف من 6 فقرات لتقرير سيطرة الطالب على وحدة تعليمية صغيرة في الرياضيات، فإن عولجت الفقرات على أنها بُعد القياس، فإن القرارات تصور على أنها تعتمد تصميم 2/ دراسة D لـ $n_i^1 = 6$ كشروط لبُعد الفقرة. وبالتالي فإن تباين الخطأ يساوي $\frac{\sigma_e^2 + \sigma_i^2}{6}$ ويمكن حساب كل من σ_i و σ_e باستخدام تحليل التباين الثنائي لإجابات الفقرة.

ويلخص جدول (8-7) اعتماد تباين الخطأ على نوع القرار، ويبين أيضاً الدرجة المستخدمة في تشكيل فترة الثقة والمعلم المحسوب لفترة الثقة لتجمعات تصميمات البعد الواحد جميعها وأنواع القرارات – مقارنة ومطلقة. إن المسوغات والمضامين لمدخلات القرارات المطلقة بسيطة نسبياً لذا فإنها عرضت أولاً. فللقرار المطلق مثلاً تستخدم الدرجة الملاحظة للتصميم المستخدم لتشكيل فترة الثقة عند جمع بيانات دراسة D وفي تقديرات فترة الثقة لدرجة النطاق. لتوضيح استخدام جدول (8-7) مع قرار مطلق افترض أن معلماً طبق اختبار نهائي

مؤلف من 100 فقرة من نوع اختيار من متعدد واستخدم نتائجه لتأشير الدرجات المناظرة للنسب المئوية للاجابات الصحيحة:

جدول (8-7): الدرجات وتباين الخطأ لبناء فترات الثقة المناسبة لتجمع كل من نوع القرار والتصميم

نوع القرار	التصميم	الدرجة	تباين الخطأ	المعلم
مطلق	1	X_{pi}	$\sigma_e^2+\sigma_i^2$	μ_p
	2	X_{pl}	$n_i^{1}/(\sigma_e^2+\sigma_i^2)$	μ_p
	3	X_{pi}	$\sigma_e^2+\sigma_i^2$	μ_p
	4	X_{pl}	$n_i^{1}/(\sigma_e^2+\sigma_i^2)$	μ_p
مقارن	1	$X_{pi}-X_{pi}$	σ_e^2	μ_p-M
	2	$X_{pl}-X_{pl}$	$n_i\sqrt{}\,\sigma_e^2$	μ_p-M
	3	X_{pi}	$\sigma_e^2+\sigma_i^2$	μp
	4	X_{pl}	$n_i\sqrt{}\,\sigma_e^2+\sigma_i^2$	μ_p

متصل التدريج:

A: 1.00 - .91

B: .90 - .81

C: .80 - .71

D: .70 - .61

F: .60 - s.00

ويعد التطبيق على المفحوصين دراسة D، تؤلف فيه الفقرات بعدُ القياس المفرد، ولأن المفحوصين جميعاً تقدموا للفقرات نفسها فإن دراسة D أجريت بتصميم 2. وتعد قرارات التدريج قرارات مطلقة إذ أن الدرجة تعطي للمفحوص دون معرفة درجات الطلبة الآخرين في الصف. وبالعودة إلى جدول (8-7) فإن تباين الخطأ يكون $n_i\sqrt{}\,\sigma_e^2+\sigma_i^2$، والدرجة المستخدمة في تكوين فترة الثقة هي X_{pl}، وتحسب فترة الثقة من درجة النطاق μ_p. وفي هذه التعبيرات

تمثل n_i عدد فقرات الاختبار، ونسبة الإجابات الصحيحة تمثلها x_{pI}. ويمكن إجراء حساب σ_e^2 وσ_i^2 باستخدام تحليل التباين الثنائي لبيانات دراسة D.

يلاحظ من جدول (8-7) أن مضمون قرارات المقارنة مباشرة نسبياً، ومع ذلك فإن مسوغات دراسة D/ تصميم 1 معقدة بعض الشيء. مع ذلك سنجري توضيحاً لدراسة D/ تصميم 1، وملاحظة أن مسوغات مشابهة تطبق على التصميمات الأخرى، وتذكر أن التصميم 1 تصميم متقاطع يتعرض فيه المفحوصين جميعاً لشرط قياسي واحد. افترض أننا نجري دراسة D باستخدام التصميم 1 وأنه بإمكاننا قياس المجتمع بأكمله، وهذا يعني أننا نعرف القيمة العددية لـ μ_i: متوسط درجة المجتمع لشرط القياس الواحد. تحت أي الشروط تؤخذ قرارات المقارنة باستخدام الدرجة الملاحظة x_{pi} التي تكون مشابهة لتلك التي يمكن اتخاذها فيما لو عرفنا درجات النطاق لكل مفحوص؟ يمكن الإجابة عن السؤال بسهولة إذا ميزنا حقائق قليلة واسترجعنا النموذج الخطي لـ x_{pi}. في البداية نتحدث عن الحقائق: تذكر أن قرار المقارنة يستخدم مقارنة درجات ملاحظة للمفحوصين، فللتصميم 1 تتكافأ مقارنة الدرجات (x_{pi}) ومقارنة الدرجات المنحرفة (x_{pi}-μ_i) إذ أن μ_i هي نفسها للمفحوصين جميعاً. وبطريقة مشابهة فإن مقارنة درجات النطاق مكافئة لمقارنة درجات النطاق المنحرفة (μ_p-μ)، وهاتان المتكافئتان هي الحقائق التي نحتاجها. والآن دعنا نسترجع النموذج الخطي لـ x_{pi}:

$$e_{pi} + (\mu-\mu_i) + (\mu-\mu p) + \mu = x_{pi} \qquad (21\text{-}8)..........$$

وبحذف μi من كلا الطرفين نحصل على:

$$e_{pi} + \mu - \mu_p = \mu_i - x_{pi} \qquad (22\text{-}8)..........$$

والمعادلة (8-22) تجعل القرار المعتمد على X_{pi}-μ مشابه للقرار المعتمد على μp-μ فيما لو كانت الدرجات المنحرفة هي نفسها. لذلك فإن الإجابة على سؤالنا هي أن القرارات المقارنة المتخذة باستخدام درجة النطاق وتزودنا بأن $\mu_i - x_{pi} = \mu - \mu_p$، ومن وجهة النظر هذه فإن الخطأ المستهدف في اتخاذ قرار المقارنة باستخدام التصميم 1 يتطلب تعويض μ_i -x_{pi} مكان $\mu - \mu_p$. وكنتيجة فإن فترة الثقة المناسبة تكون $(\mu_i - x_{pi})$ ±σ_e^2 والتي تقدر μ -μ_p. ومع ذلك فإننا لا نعرف قيمة μ_i، ولذلك فإننا نعوض بمتوسط العينة والذي نرمز له بـ x_{pi} في جدول (7-8)، والمتوسط μ_i تشير P إلى متوسط العينة عبر جميع الأشخاص.

لتوضيح استخدام جدول (8-7) مع قرارات المقارنة دعنا نعود إلى مثال عالم النفس الاجتماعي، وتذكر أن عالم النفس مهتماً بالارتباط مع الحذر، ويتطلب حساب الارتباط قرارات مقارنة لان الارتباط يقيس إلى أي مدى ارتباط الفروق الفردية على متغيرين. افترض أن عالم النفس قرر استخدام التصميم 2/ دراسة D. وبالنسبة لجدول (8-7) فإن تباين

الخطأ $\sigma_e^2 \sqrt{} n_i$، والدرجة المستخدمة في حساب فترة الثقة هي $x_{pI} - x_{pI}$، وتقدير فترة الثقة يكون $\mu_p - \mu$، وفي هذه التعبيرات تمثل n_1 عدد المقدرين في دراسة D، و x_{pI} تمثل متوسط تقديرات المفحوص p و x_{pI} تمثل متوسط x_{pi}.

ومن المهم تذكر أن الدرجة الملاحظة في نظرية إمكانية التعميم عبارة عن المتوسط عبر شروط القياس جميعها التي تعرض لها المفحوص p. ومن المألوف استخدام الدرجة الكلية المناظرة بدلاً من متوسط الدرجة في صنع القرار. وفي مثال الاختبار محكي المرجع يظهر هذا إذا استخدمت $X_{pI}\ n_i^2 = Y$. ولمواءمة استخدام الدرجة الكلية أضرب تباين درجة الخطأ بمربع عدد الشروط التي قيس المفحوص فيها. وفي مثال الاختبار محكي المرجع فإن تباين الدرجة الخطأ لـ y:

$$(\sigma_e^2 + \sigma_i^2)\ n_i = [\ n_i \sqrt{}\ (\sigma_e^2 + \sigma_i^2)\ ]\ n_i \qquad (23 - 8)$$

نظرية التعميم للتصميمات ثنائية البعد:

في العديد من المواقف يجب تصنيف شروط القياس إلى بعدين أو أكثر. ومثال لدراسة بُعدين يظهر عندما يستجيب المفحوصين على عدة أسئلة مقالية وتصحيح من قبل عدة مقدرين، ويؤلف كل من مجموعة المقدرين ومجموعةالأسئلة المقالية شروط القياس (الأبعاد). وعندما تصنف شروط القياس إلى بُعدين فإن النظرية والطرائق المطورة للتصميمات أحادية البعد تكون غير مناسبة لصياغة معامل إمكانية التعميم وتقديره. وفي هذا الجزء سنعرض صياغة وتقدير معاملات الثبات للتصميمات ثنائية البُعد. ويتضمن العرض الموضوعات الآتية:

1/ المصطلحات لوصف تصميمات التعميم ونطاقاته.

2/ تباين درجة النطاق لنطاقات تعميم مختلفة.

3/ تباين الدرجة المتوقع لتصميمات دراسة D.

4/ صياغة معاملات إمكانية التعميم.

5/ تقدير معاملات إمكانية التعميم.

المصطلحات:

يقال على البعدين أنهما تقاطعا (Crossed) عندما يظهر كل شرط قياس في البعد الأول مجتمعاً مع كل شرط قياس للبعد الثاني. على سبيل المثال: عند تطبيق كل اختبار من مجموعة اختبارات في المواقف العديدة جميعها، ونقول أن بعدي الاختبارات والمواقف تقاطعت. ويقال على البعد أنه تداخل nested مع البعد الآخر إذا ظهرت شروط قياس البعد الأول المختلفة

مرتبطة مع شروط قياس البعد الثاني جميعها. على سبيل المثال افترض أن كل تلميذ من المئة تلميذ أجابوا على سؤال مقالي واحد في الفيزياء، وواحد في الرواية الأمريكية والأخير في تاريخ الرياضيات، وحيث أن الموضوعات مختلفة فإن مجموعة مقدرين (المصححين) تصحح سؤال الفيزياء، وثانية لسؤال الرواية وثالثة لسؤال التاريخ. والمجموعة الأولى من المقدرين صححت سؤال الفيزياء للمفحوصين جميعاً، وهكذا... هنا نقول أن المقدرين تداخلوا ضمن الأسئلة المقالية. لاحظ أنه بالتصميمات أحادية البعد يمكن لبعد القياس أن يتقاطع أو يتداخل مع المفحوصين. وفي التصميمات ثنائية البعد فإن أي من البعدين يمكن أن يتقاطع أو يتداخل مع المفحوصين. وفي المثال الحالي فإن كل من المقدرين والموضوعات تقاطعت مع المفحوصين إذ أن كل مفحوص كتب مقالة (أجاب عن سؤال مقالي) في كل موضوع وقدِّر من قبل المقدرين جميعاً.

تباين الدرجة الشاملة:

من الضروري قبل وصف نطاقات التعميم وتباين الدرجة الشاملة المناظر تحديد عناصر التباين المستخدم في نظرية إمكانية التعميم للتصميمات ثنائية البعد. وهذه العناصر هي:

1- $\sigma_p{}^2$ تباين المفحوص.

2- $\sigma_i{}^2$ تباين شروط العامل الأول i.

3- $\sigma_j{}^2$ تباين شرط العامل j.

4- $\sigma_p{}^2{}_i$ تباين التفاعل بين المفحوصين والعامل i.

5- $\sigma_p{}^2{}_j$ تباين التفاعل بين المفحوصين والعامل j.

6- $\sigma_i{}^2{}_j$ تباين التفاعل بين المفحوصين والعاملين في i و j.

7- $\sigma_r{}^2$ تباين البواقي.

ويعتمد تباين الدرجة الشاملة على الأبعاد (العوامل) فإن كان أحدها ثابتاً في نطاق التعميم سيكون هنالك خمسة نطاقات تعميم ممكنة وتباين درجة شاملة لكل منها ومع ذلك فمن المهم الإشارة إلى أنه خلال مناقشتنا لنظرية إمكانية التعميم لبعدين نفترض أنه تم تحديد نطاق تعميم يتقاطع مع أبعاد القياس (ولا نفترض تقاطع أبعاد القياس مع بعضها البعض في التصميمات التي تستخدم دراسات G أو D. ومن الممكن استخدام تصميم ذو شرط قياسي واحد متداخل مع شرط قياسي آخر، مع أن الأبعاد تقاطعت في نطاق التعميم)، وقد تم وضع هذا الافتراض للحفاظ على النقاش من زيادة التعقيد. هذا ويمكن لنظرية إمكانية التعميم أن

تتناول نطاقات ببُعد واحد تتداخل مع غيرها. وقد ناقش كرونباخ وزملاؤه (Cronbach etal., 1972) هذا النوع من النطاقات. النطاق 1 : يفترض في النطاق الأول أن البعدين عشوائيين، ومثال على هذا النطاق تعميم الفاحص عبر صيغ اختبارية عديدة ومواقف قياس، وكل منهما كان بعداً عشوائياً. وترمز σ_p^2 إلى تباين درجة النطاق، وفي هذه الحالة (الأبعاد عشوائية) فإن تباين الدرجة الشاملة لا يعتمد على كون تقاطع البعد أو تداخله مع المفحوصين.

النطاقين 2 و 3 : في هذين النطاقين يكون أحد البعدين ثابت والآخر عشوائي، مع تقاطع العامل الثابت مع المفحوصين. ففي النطاق، 2 يكون i البعد الثابت ودرجة تباين النطاق يكون $n_i\sqrt{} \sigma^2_{pi}+ \sigma_p^2$. والمثال في النطاق 1 يخدم هنا إذا افترضنا أن المواقف عشوائية والصيغ المرمز لها بالبعد i كانت ثابتة. وفي النطاق 3 يكون البعد j ثابتاً، لذا فإن تباين درجة النطاق يكون: $n_i\sqrt{} \sigma^2_{pj}+ \sigma_p^2$ ونحصل على النطاق 3 عندما تكون المواقف ثابتة والصيغ عشوائية، وتشير الرموز n_{ij} و $n_i\backslash$ إلى عدد الشروط للأبعاد i و j في دراسة D.

النطاقين 4 ، 5 : وفي هذين النطاقين يكون أحد البعدين ثابت والآخر عشوائي مع تداخل البعد الثابت مع المفحوصين. كمثال عند تدريج الآباء لقدرات أبناءهم في تناول سلسلة مواقف ضغط، وحيث أن لكل طفل والدين أثنين فقط وهؤلاء مختلفين لكل طفل فإن بعد الآباء يكون ثابتاً مع تداخل الآباء مع الأطفال، ويُعد بُعد الموقف عشوائي. ففي النطاق 4 تكون I هي البعد الثابت ويوجد عدد شروط $n_i\backslash$ لـ I تتداخل مع كل مفحوص، ويكون تباين درجة النطاق $n_i\sqrt{} \sigma^2_{pi}+ \sigma_i^2 + \sigma^2_p$ وفي النطاق 5 يكون j هو البعد الثابت وتباين النطاق يكون $n_i\sqrt{} \sigma^2_{pj}+ \sigma_j^2 + \sigma^2_p$.

وعندما نحصل على تباين درجة النطاق لنطاقات التعميم جميعها، يمكننا عرض تباين الدرجة الملاحظة المتوقعة للعديد من تصميمات دراسة D. وكما سنرى في الجزء التالي فإن معامل إمكانية التعميم يعبر عنه بنسبة تباين النطاق إلى تباين الدرجة الملاحظة، والذي يعتمد على نطاق التعميم وتصميم دراسة D.

تباين الدرجة الملاحظة المتوقع:

يعتمد تحديد تباين الدرجة الملاحظة المتوقع على تصميم دراسة D. وسنعرض خمسة تصميمات ممكنة لدراسة D، وقد عرض كل من جليسر وكرونباخ وراجاراتنام و (Gleser, Cronbach & Rajaratnam, 1965) ثلاثة تصميمات أخرى.

تصميم 1: هنا تتقاطع العوامل مع بعضها البعض ومع المفحوصين، ويظهر كل شرط للبعد I مجتمعاً مع كل شرط للبعد J، ويتعرض المفحوص لتجمعات الشروط جميعها. كمثال فعند

تطبيق اختبارات عديدة (بُعد I) في الموقف نفسه ويتقدم المفحوصين جميعاً إلى الاختبارات جميعها وفي الشروط جميعها. وقدم كرونباخ الرسم $p(x_i\ x_j)$ لهذا التصميم، وترمز p إلى المفحوصين و i إلى شروط البعد I، و j إلى شروط البعد J. لذا فإن $p(x_i\ ,\ x_j)$ تؤشر إلى تقاطع المفحوصين وشروط البُعدين I و J. وفي التصميم i الدرجة الملاحظة تساوي متوسط الملاحظات $n_i\backslash n_j\backslash$، وتباين الدرجة المتوقع يكون:

$$\sigma_e^2\ \frac{1}{n_i\backslash n_j\backslash}\ +\sigma^2{pj}\ \frac{1}{n_j\backslash}\ +\sigma^2{pi}\ \frac{1}{n_1\backslash}\ +\sigma^2p \qquad (24\text{ -}8)\ ....$$

تصميم 2: في هذا التصميم تتداخل شروط $n_j\backslash$ للبعد J مع شروط $n_i\backslash$ للبعد I، وكلا البعدين يتقاطع مع المفحوصين مثال ذلك: تمثل ثلاثة أسئلة مقالية البعد I، صحح كل منها من قبل ثلاثة مصححين مختلفين (بعد J)، وتداخل كل مصحح، مع كل سؤال مقالي، وكل مفحوص أجاب على الأسئلة المقالية الثلاثة وبالتالي تم إعطاءه درجة من قبل كل مصحح، لذلك فإن كل مفحوص تقاطع مع كل سـؤال وكل مصحح، ويرمز إلى هذا التصميم بالرمز xp (j : i)، وتقرأ شروط J متداخلة مع شروط البعد I وتجمع الشروط تقاطع مع المفحوصين. والدرجة الملاحظة تكون متوسط الملاحظات $n_i\backslash n_j\backslash$، وتباين الدرجة الملاحظة هو:

$$\sigma_e{}^2 + \sigma^2{}_{pj}\ \frac{1}{n_i\backslash n_j\backslash}\ +\sigma^2{pi}\ \frac{1}{n_i\backslash}\ +\sigma^2p \qquad (25\text{ -}8)\ ....$$

وفي التصميم 2 تتداخل شروط I مع شروط J (وعلى القاريء ملاحظة أنه يمكن استخدام تصميمات معينة لدراسات G لحساب تقديرات منفصلة لكل من $\sigma^2{}_{pj}$ و $\sigma_e{}^2$، في حين أنه من الضروري في التصميمات الأخرى حساب تقدير واحد لـ $\sigma_e{}^2 + \sigma^2{}_{pj}$، وينطبق هذا على عناصر التباين جميعها الموجودة بين قوسين في المعادلات من (8-25) إلى (8-28). وكل معادلة منها تعبر عن تباين الدرجة الملاحظ.

تصميم 3: هنا يظهر كل شرط لـ I مجتمعاً مع أحد شروط J، لذا فإنه يوجد عدد شروط كلي nj ni = k. ويتعرض كل مفحوص لشروط القياس k جميعها وبالتالي فإن كل بُعد يتقاطع مع المفحوصين، ويرمز لهذا التصميم بالرمز xp(ι,j) . وفي مثال الأسئلة المقالية-التصحيح يظهر هذا التصميم فيما لو كان هنالك مصححاً واحداً لسؤال الفيزياء وآخر لسؤال الرواية وآخر لسؤال الرياضيات. وتكون الدرجة الملاحظة عبارة عن متوسط الشروط k، وتباين الدرجة الملاحظ يكون:

$$(\sigma_e{}^2 + \sigma^2{}_{pj} + \sigma^2{}_{pi})\ \frac{1}{k}\ +\sigma^2p \qquad (26\text{ -}8)\ ....$$

تصميم 4: تتقاطع هنا شروط البعد I ($n_i^{\backslash}$) مع المفحوصين، في حين تتداخل شروط $n_j^{\backslash}$ للبعد J مع المفحوصين، مع تقاطع البعدين I و J. ومثال ذلك عندما يتم تدريج المعلمين من قبل طلبتهم في مساقات عديدة وعلى متصلات التدريج نفسه، ويعد المعلمون هنا هم المفحوصين ويؤلف التلاميذ شروط البعد J. وإذا افترضنا أنه لا تلاميذ انخرطوا في أكثر من مساق، وهكذا يتداخل المفحوصين مع التلاميذ، وبما أن التلاميذ جميعهم استخدموا التدريج نفسه (البعد I)، والمفحوصين جميعاً تم تدريجهم على هذا التدريج فإن شروط البعد J تقاطعت مع شروط البعد I والمفحوصين، ويرمز لهذا التصميم بـ $X_t(j : p)$. وقد ناقش كل من كين وجيلمور وكروكس (Kane, Gillmorem Crooks, 1976) هذا المثال بالتفصيل. وفي هذا التصميم تباين الدرجة الملاحظة المتوقع يكون:

$$(\sigma_e^2+\sigma^2_{ij})\ \frac{1}{n_i^{\backslash}\, n_j^{\backslash}} + (\sigma_j^2+\sigma^2_{pj})\ \frac{1}{n_j^{\backslash}} + \sigma^2_{pi}\ \frac{1}{n_i^{\backslash}} + \sigma^2_p \qquad \text{........ (27 -8)}$$

وهناك تصميم 4 حيث تتقاطع شروط J مع المفحوصين في حين تتداخل شروط I مع المفحوصين.

تصميم 5: هنا تتقاطع I مع المفحوصين في حين تتداخل J مع المفحوصين، بالإضافة إلى تداخل J مع I، ويرمز لهذا التصميم بـ: J:(ixp)، ومثال على هذا التصميم الدراسة التي تستهدف قياس الوقت المستغرق في المهمة، إذ سيتم ملاحظة الطلبة لفترات كل منها 15 دقيقة، وكل التلاميذ تمت ملاحظتهم من قبل الملاحظ نفسه $n_i^{\backslash}$، لذا فإن الملاحظين تقاطعوا مع المفحوصين، وكل تلميذ تمت ملاحظته لفترات عددها $n_j^{\backslash}$ من قبل الملاحظين جميعهم، والفترات $n_j^{\backslash}$ مختلفة لكل ملاحظ وتداخلت الفترات (بعد J) مع الملاحظين (بعد I)، علاوة على أنه لا يوجد تلميذين تمت ملاحظتهما في الفترة الزمنية نفسها، لذا فإن الفترات تداخلت ضمن المفحوصين. وتباين الدرجة الملاحظة المتوقع لهذا التصميم هو:

$$(\sigma_e^2+\sigma^2_{ij}+\sigma^2_j+\sigma^2_{pj})\ \frac{1}{\hat{n}_i\, \hat{n}_j} + \sigma^2_{pi}\ \frac{1}{n_i^{\backslash}} + \sigma^2_p \qquad \text{........ (28 -8)}$$

هذا ويوجد تصميم $5^{\backslash}$ حيث تتداخل J مع المفحوصين، وتتقاطع I مع المفحوصين وتتداخل I في J.

معاملات إمكانية التعميم:

إن صياغة معامل إمكانية التعميم المناسب يعتمد على نطاق التعميم وتصميم دراسة D. وحالما يتم تحديد النطاق وتصميم دراسة D. وحالما يتم تحديد النطاق وتصميم دراسة D فإن

معامل التعميم ببساطة يكون نسبة تباين النطاق إلى تباين الدرجة الملاحظ. وكمثال لنعد إلى أسئلة المقال الثلاثة.

فيزياء، ورواية، ورياضيات التي صححت من قبل مصححين عدة، وافترض أن نطاق التعميم حدد بمعالجة الأسئلة (البعد I) على أنها بعد ثالث والمصححين على أنهم بعد عشوائي، هنا يكون نطاق التعميم هو تصميم 2، وإذا أجريت الدراسة D بتصميم 2 فإن معامل إمكانية التعميم المناسب يكون:

$$\frac{n_i' / \sigma^2_{pi} + \sigma_p^2}{n_i' \, n_j' / (\sigma_e^2 + \sigma^2_{pj}) + n_i' / \sigma^2_{pi} + \sigma_p^2} = \rho^2$$

اختيار تصميم دراسة G: كمبدأ يجب أن تحسب معاملات إمكانية التعميم من البيانات المتجمعة من دراسة G والتي تجري قبل دراسة D. وهذا يسمح للباحث الحكم على ما إذا كانت بيانات دراسة D لها إمكانية تعميم مناسبة. وفي تصميم دراسة G يجب الموازنة بين قضايا ثلاث:

الأولى: يجب أن يكون تصميم دراسة G منسجماً مع نطاق التعميم الذي يفكر الباحث فيه. مثال ذلك إذا أراد الباحث اعتبار صيغ الاختبار (بعد I) بعداً عشوائياً ومواقف القياس (بعد J) بعداً ثابتاً في نطاق التعميم، فإن تقاطعت المواقف مع المفحوصين في النطاق فإن التصميم يكون نطاق 3، وضمن هذا النطاق فإن استخدام تصميم 4 بوجود تداخل بين المفحوصين والمواقف لا يؤثر على شيء. والمبدأ هنا أن البعد الثابت الذي تقاطع مع المفحوصين في نطاق التعميم يجب أن يتداخل مع المفحوصين في تصميم دراسة G.

الثانية: إذا كان التصميم يسمح بتقدير عناصر التباين لتباين درجات النطاق، فإن كان تباين درجات النطاق يتضمن تباين عنصر معين، وفي الوقت نفسه لا يسمح التصميم بتقدير هذا العنصر فلا حرج من اعتبار التصميم على أنه تصميم دراسة G ممكن. وقد تم إعداد جدول (8-8) ليسمح للقارئ بتناول هذه القضايا، فهو يقدم التصميمات المناسبة لدراسة G لكل نطاق تعميم. وباستخدام جدول (8-8) لمثال الاختبار والمواقف نرى أنه وللنطاق 3 تكون التصميمات (5,4,2,1) مفيدة في تصميمات دراسة G.

الثالثة: عدم تجاهل حقيقة أن البيانات المتجمعة لتصميم معين في دراسة G لا يمكن استخدامها لتقدير حساب تباين الدرجة الملاحظة لتصميمات دراسة D جميعها. ويكمن الهدف في اختيار تصميم دراسة G التي يمكن استخدامها في حساب تباينات الدرجة الملاحظة لأكبر عدد ممكن من تصميمات دراسة D. ويؤشر الجدول (8-9) تصميمات دراسة G التي يمكن استخدامها .

جدول (8-8): تصميمات دراسة G لكل نطاق تعميم

وصف البعد					تصميمات دراسة G							
نطاق التعميم	عشوائي	ثابت	تقاطع مع p	تداخل مع P*	i	2	2\	3	4	4\	5	5\
					pxi xj	(j:i) xp	(:J)xp	(i, J) xp	(j:p)xi	(i:p)xj	j:(ixp)	i: (jxp)
1	L.J	-	-	-	*	*	*	*	*	*	*	*
2	J	L	L	-	*	*			*		*	
3	L	J	J	-	*		*			*		*
4	J	L	-	L						*		
5	L	J	-	J					*			

* تشير إلى أبعاد ثابتة، تتقاطع أو تتداخل في p لا تؤدي إلى مضامين لاختيار تصميم دراسة G

جدول (8-9): تصميمات دراسة G المحتملة لحساب تباين الدرجة الملاحظة المتوقع في دراسات D.

تصميمات دراسة G		تصميمات دراسة D							
		1	2	/2	3	4	/4	5	/5
1	(Pxi xj)	*	*	*	*	*	*	*	*
2	xp(j : i)		*					*	
2\	xp(i : j)			*					*
3	xp(i : P)				*				
4	xi(j : p)					*		*	
4\	xj (i : p)						*		*
5	j : (ixp)							*	
5\	i : (j xp)								*

لحساب تباين الدرجة الملاحظة لتصميمات دراسة D جميعها. لذا فإن كان تصميم 1 عملي فيجب استخدامه في دراسة G. لاحظ أن التصميم 1يسمح بحساب تباينات الدرجة الملاحظ لتصميمات D الثمانية الممكنة. وفي اختيار تصميم دراسة D لمثال الصيغ والمواقف (نطاق 3) يجب إهمال التصميمات (5،4،3،2)، وذلك لأنها لا تغير في تصميمات دراسة G، وهي كذلك لتصميمات دراسة D في المثال.

حساب معاملات إمكانية التعميم:

حالما يتم إجراء دراسة G فإن تحليل التباين المناسب للبيانات يؤدي إلى حساب عناصر التباين المطلوبة لحساب معامل إمكانية التعميم. وهناك محددات تمنع إجراء المعالجة التامة لتحليل التباين للتصميمات كلها. ومع ذلك فإن عوامل كل منها مبينة في جدول (8-10)، بالإضافة إلى متوسط المربعات المتوقع لكل عامل منها. وتم صياغة هذه المتوسطات من خلال عناصر التباين التي حددت فيما سبق. والقراء الذين ليس لديهم معرفة بتحليل التباين قد يجدوا بعض متوسطات المربعات في التصميمات من (2 إلى 5) مربكة في البداية. مثلاً خذ مصدر تباين J : I في تصميم 2، وباستخدام النتائج التجريبية التي قدمها ميلمان وجلاس

(Millman&Glass,1967) يكون متوسط المربعات المتوقع على النحوالآتي: $np\sigma_{ji}^2 + \sigma p_j^2 = \sigma p_j^2$ ، والحد الأخير هنا عبارة عن تباين شروط j المتداخلة مع شروط i، ومع ذلك يمكن تبيان أن $\sigma_{ij}^2 + \sigma_j^2 = \sigma^2_{j.i}$ والجانب الأيسر من المعادلة مستخدم في جدول (8-10)، افترض أن حساب P^2 كما حدد في المعادلة (8-29)، فإن عناصر التباين الضرورية هي σ_p^2 و σ^2_{pi} و $(\sigma_e^2 + \sigma^2_{pj})$ واعتماداً على EMS المدونة في جدول (8-10) للتصميم 2، فإن عناصر التباين تحسب على النحو الآتي:

$$\hat{\sigma}_p^2 = \frac{MS_{pi} - MS_p}{n_i\, n_j} \quad \text{30–أ} \ldots\ldots\ldots\ldots$$

$$\hat{\sigma}^2_{pi} = \frac{MS_r - MS_{pi}}{n_i} \quad \text{30–ب} \ldots\ldots\ldots\ldots$$

$$(MS_r) = \hat{\sigma}_e^2 + \hat{\sigma}^2_{pj} \quad \text{30–جـ} \ldots\ldots\ldots\ldots$$

ويحسب معامل إمكانية التعميم باستخدام المعادلة:

$$\sigma_p^2 = \frac{n_i \backslash\, \hat{\sigma}^2_{pi} + \hat{\sigma}_p^2}{n_j \backslash\, n_i \backslash\, \hat{\sigma}_e^2 + \hat{\sigma}^2_{pj} + n_i \backslash\, \hat{\sigma}^2_{pi} + \hat{\sigma}_p^2} \quad (31 - 8) \ldots\ldots\ldots\ldots$$

جدول (8-10): متوسط المربعات المتوقع للتصميمات (من 1 إلى 5)

التصميم 1

SV	EMS
P	$\sigma_e^2 + n_i\, n_j \sigma_p^2 + n_j \sigma^2_{pi} + n_i\, \sigma^2_{pj}$
L	$\sigma_e^2 + n_j\, n_p \sigma_i^2 + n_j\, \sigma^2_{pi} + n_p \sigma^2_{ij}$
J	$\sigma_e^2 + n_i\, n_p \sigma_j^2 + n_i\, \sigma^2_{pj} + n_p \sigma^2_{ij}$
pl	$\sigma_e^2 + n_i\, \sigma^2_{pi}$
pj	$\sigma_e^2 + n_i\, \sigma^2_{pj}$
IJ	$\sigma_e^2 + n_p\, \sigma^2_{ij}$
البواقي	σ_e^2

التصميم 2

SV	EMS
P	$\sigma_e^2 + \sigma^2_{pj} + n_i n_j \sigma_p^2 + n_j \sigma^2_{pi}$
L	$\sigma_e^2 + \sigma^2_{pj} + n_p n_j \sigma_i^2 + n_j \sigma^2_{pi} + n_i (\sigma^2_j + \sigma^2_{ij})$
j:1	$\sigma_e^2 + \sigma^2_{pj} + n_i (\sigma^2_j + \sigma^2_{ij})$
pL	$\sigma_e^2 + \sigma_p^2 + n_j \sigma^2_{pi}$
البواقي	$\sigma_e^2 + \sigma^2_{pj}$

التصميم 3

SV	EMS
P	$\sigma_e^2 + \sigma^2_{pj} + \sigma^2_{pi} + k \sigma^2_p$
L : j	$\sigma_e^2 + \sigma^2_{pj} + \sigma^2_{pi} + n_p (\sigma^2_i + \sigma^2_j + \sigma^2_{ij})$
البواقي	$\sigma_e^2 + \sigma^2_{pj} + \sigma^2_{pi}$

التصميم 4

SV	EMS
P	$\sigma_e^2 + \sigma^2_{ij} + n_i n_j \sigma_p^2 + n_j \sigma^2_{pi} + n_i (\sigma^2_j + \sigma^2_{pl})$
L	$\sigma_e^2 + \sigma^2_{ij} + n_p n_j \sigma_i^2 + n_i \sigma^2_{pi}$
pl	$\sigma_e^2 + \sigma^2_{ij} + n_j \sigma^2_{pi}$
j : p	$\sigma_e^2 + \sigma^2_{ij} + n_i (\sigma^2_j + \sigma^2_{pj})$
البواقي	$\sigma_e^2 + \sigma^2_{ij}$

التصميم 5

SV	EMS
P	$\sigma_e^2 + \sigma^2_j + \sigma^2_{ij} + \sigma^2_{pj} + n_i n_j \sigma^2_p + n_i \sigma^2_{pi}$
L	$\sigma_e^2 + \sigma^2_j + \sigma^2_{ij} + \sigma^2_{pj} + n_i n_p \sigma^2_i + n_j \sigma^2_{pi}$
Pl	$\sigma_e^2 + \sigma^2_j + \sigma^2_{ij} + \sigma^2_{pj} + n_j \sigma^2_{pi}$
البواقي	$\sigma_e^2 + \sigma^2_j + \sigma^2_{ij} + \sigma^2_{pj}$

مثال:

لنأخذ اختبار مقالي يكتب فيه المفحوصين مقالات في مواضيع ستة. وفي دراسة إمكانية التعميم لدرجات هذا الاختبار فإن الفاحص سيعالج كلاً من المواضيع والمصححين على أنها أبعاد عشوائية، وهذا عبارة عن تصميم 1، لذا فإن تباين درجة النطاق هو σ_p^2، وبالعودة إلى جدول (8-8) نجد أن استخدام التصميم 1 في دراسة G سيسمح بتقدير تباين الدرجة الملاحظ لأكبر قدر من تنوع تصميمات دراسة D. ولنفترض أن استخدام التصميم قابل للتطبيق كدراسة G، لذا فإن الباحث سيجري دراسة G باستخدام هذا التصميم، وقد تم اختيار سبعة مفحوصين كتبوا مقالات في موضوعين وقد صممت المقالات بوساطة ثلاثة مصححين. ويبين جدول (8-11) خلاصة تحليل التباين لبيانات دراسة G.

كيف يمكننا استخدام بيانات جدول (8-11) في اختيار تصميم دراسة D في جمع بيانات المفحوصين. افترض أن الباحث وجد أن استخدام ستة مصححين عملياً في دراسة D، ومع هذا العدد من المصححيين فإن أحد تصميمات دراسة D الممكنة هو التصميم 1. وبإعطاء بعد المصححيين الرمز J، وبعد الموضوعات الرمز I، نحصل على دراسة D بستة مصححين ($n_j\backslash = 6$) وستة موضوعات ($n_i\backslash = 6$). ويكون تباين الدرجة الملاحظ للتصميم 1 هو:

$$n_i\backslash\, n_p / \sigma_e^2 + n_j\backslash / \sigma^2_{pj} + n_i / \sigma^2_{pi} + \sigma^2\rho \quad \text{.......... (8- 32)}$$

وكنتيجة، فإن معامل إمكانية التعميم المناسب هو:

$$\rho^2 = \frac{\sigma^2_p}{n_i\backslash\, n_j / \sigma_e^2 + n_j / \sigma^2_{pj} + n_i / \sigma^2_{pi} + \sigma^2_p} \quad \text{.......... (8- 33)}$$

ويعد التصميم 3 هنا تصميماً ممكناً، إذ يصحح كل مصحح موضوعاً واحداً فقط، لذا فإن $K = n_i\backslash = n_j\backslash = 6$.

وتباين الدرجة الملاحظ يكون:

$$K/\ (\sigma_e^2 + \sigma^2\rho_j + \sigma^2\rho_i) + \sigma^2\rho \quad \text{.......... (8- 34)}$$

وبالتالي فإن معامل إمكانية التعميم المناسب هو:

$$\rho^2 = \frac{\sigma^2_p}{K/\ (\sigma_e^2 + \sigma^2_{pj} + \sigma^2_{pi}) + \sigma^2_p} \quad \text{.......... (8- 35)}$$

والخطوة التالية هي تطوير معادلات لحساب عناصر التباين الضرورية لكل معامل إمكانية تعميم. ولحساب كل من تباين الدرجة الملاحظ وتباين النطاق يجب أن نحسب σ_p^2، σ_{pi}^2. وباستخدام مربعات المتوسطات المتوقع. لتصميم 1 في جدول (8-10) نحصل على المعادلات الآتية:

جدول (8-11): خلاصة تحليل التباين لتصميم ثنائي البعد:

MS	df	SS	SV
11.095	6	66.571	الأفراد (P)
.857	1	.857	الموضوعات (L)
3.714	2	7.428	المصححين (J)
1.301	6	7.809	PL
4.547	12	54.571	pj
1.142	2	2.285	Lj
1.087	12	152.571	البواقي

$$\hat{\sigma}\rho^2 = \frac{(MS_r + MS_{pj} - MS_{pi} - MS_p)}{n_i\, n_j}$$ .... (8- 36 أ)

$$\hat{\sigma}P_{pi}^2 = \frac{MS_r - MS_{pi}}{n_j}$$ .... (8- 36 ب)

$$\hat{\sigma}P_{pj}^2 = \frac{MS_r - MS_{pj}}{n_i}$$ .... (8- 36 جـ)

$$\hat{\sigma}e^2 = MS_r$$ .... (8- 36 د)

وبالتعويض المناسب باستخدام بيانات جدول (8-11) في معادلات (8-36) نحصل على:

$$\sigma_p^2 = \frac{(1.087 + 4.547 - 1.301 - 11.095)}{6} = 1.055$$

$$\hat{\sigma}\rho^2_{pi} = \frac{(1.087 - 1.301)}{6} = 0.071$$

$$\hat{\sigma}\rho^2_{pt} = \frac{(1.087 - 4.547)}{2} = 1.730$$

و $\hat{\sigma}_e^2 = 1.087$

وبتعويض هذه التقديرات في المعادلة 8-33 نحصل على:

$$\hat{\rho}^2 = \frac{1.055}{36/1.037 + 6/1.730 + 6/.071+1.055} = 0.76$$

وكما حسبنا معامل إمكانية التعميم للتصميم 1، نحصل على معامل إمكانية التعميم للتصميم 3 بالتعويض في المعادلة (8-35) وعلى النحو الآتي:

$$\hat{\rho}^2 = \frac{1.055}{6/(1.087 + 1.730 + .071) +1.055} = 0.69$$

وبما أن كلاً التصميمين أديا إلى معاملات إمكانية تعميم ليست مختلفة كثيراً، وحيث أن تصميم3 يتطلب وقتاً أقل من كل مصحح، يمكننا تقديم النصح للفاحص باستخدام تصميم 3.

افترض الآن أن الفاحص شعر أن كلا التصميمين أديا إلى معامل إمكانية تعميم مناسب، مع ملاحظة الفاحص أن عنصر التباين المساهم بأكبر قدر في تباين الدرجة الملاحظ هو σ^2_{pj} والذين يعزى إلى تفاعل المفحوص والمصحح. ويمكن تقليل أثر هذا العنصر بزيادة عدد المصححين. فمثلاً يمكن استخدام مصححين أثنين لتصحيح كل مقالة في دراسة D وبتداخل المصححين مع الموضوعات، وكل منهما يتقاطع مع المفحوصين. ويعد هذا مثال لتصميم 2 بـ 6 $n_t' = n_j' = 2$ و (لاحظ أنه في التصميم 2 أن n_j' ليست عدد المصححين الكلي، ولكنه عدد المصححين لكل مقالة). ويكون معامل إمكانية التعميم المناسب:

$$\hat{\rho}^2 = \frac{\sigma^2_p}{n_i' n_j' / (\hat{\sigma}_e^2 + \hat{\sigma}_{pj}^2) + n_i' / \sigma^2_{pi} + \sigma^2_p} \qquad (37\text{-}8) \ldots\ldots\ldots\ldots$$

وفي مثالنا نحصل على:

$$0.81 = \frac{1.055}{12/(1.087 + 1.730) + 6/.071 + 1.055} = \hat{p}$$

لذا فإن تصميم 2 (وبمصححين أثنين لكل موضوع) يزيد من معامل إمكانية التعميم فعلياً أكثر من تصميم 3 (مصحح واحد لكل مقالة)، وهذا يؤدي إلى زيادة متوسطة عن التصميم 1. ومع ذلك فإن التصميم 2 يتطلب 12 مصحح الذي قد يكون غير ممكن عملياً، وإن كان ممكناً فمن الأفضل استخدام تصميم 3 وزيادة كلاً من عدد الموضوعات والمصححين.

الخلاصة:

تهتم نظرية أمكانية التعميم بمجموعة تقنيات تستخدم لدراسة مدى تعميم قياسات مفحوص ما على مجموعة قياسات أوسع للمفحوص نفسه. وتعد مجموعة القياسات الأولى عينة ممثلة للمجموعة الثانية وقياسات المجموعة الثانية هي مجموعة افتراضية. ويتم إجراء كلاً من مجموعتي القياس ضمن شروط قياسية معينة.ويشير مصطلح الشروط هنا إلى مظاهر عملية القياس مثل المهمة، ووقت القياس، والاتجاهات، وهكذا. ويطلق على ظروف القياس لمجموعة القياسات الثانية اسم «نطاق التعميم».

ويمكن تصنيف ظروف القياس لكل مجموعة قياسات على أنها أبعاد. فعلى سبيل المثال إذا ما حددت ظروف القياس من خلال موضوع المقالة وهوية المقدِّر فإنه يتوافر في هذه الحالة بُعدين هما:

الموضوع والمقدِّر. وقد يكون البعد ثابت أو عشوائي، فإن كان البُعد يشير إلى ظروف نطاق التعميم بالضبط ولمجموعة القياسات التي يتم الحصول عليها فإن البعد يكون ثابت. وإن كان البُعد يؤشر إلى مجموعة أكثر شمولاً من ظروف نطاق التعميم فإن البعد يكون عشوائي. ويمكن للبعدين أن يتقاطع كل منهما مع الآخر، وكذلك يمكن لأحدهما أن يتداخل مع الآخر إن كان من الممكن ظهور ظروف البُعد الأول مع واحد فقط من ظروف البعد الثاني. ويمكن للبعد أن يتقاطع أو يتداخل مع المفحوصين. ويستخدم تعريف نطاق التعميم، ويمكن تحديد الأبعاد، وبيان تصنيف البعد على أنه ثابت أو عشوائي، ووصف التقاطع أو التداخل بين الأبعاد، وبيان ما إذا كان كل بعد يتقاطع أو يتداخل مع المفحوصين.

وينتج عن العملية الافتراضية لقياس المفحوص تحت الشروط جميعها في نطاق التعميم وحساب متوسط درجة نطاق المفحوص. ويمكن تكميم درجة تعميم القياس الذي نحصل عليه على درجات النطاق لمجموعة مفحوصين بوساطة معامل إمكانية التعميم أو تباين الخطأ. وأكثر أنواع معاملات إمكانية التعميم شيوعاً وانتشاراً هو نسبة تباين درجة النطاق إلى تباين الدرجة الملاحظة. ويحدد تباين درجة النطاق من خلال نطاق التعميم، ويحدد تباين الدرجة الملاحظة من خلال التصميم التي جمعت بوساطته مجموعة القياسات. ويعتمد تباين الخطأ على استخدام الدرجات لما وضعت له كما هو لنطاق التعميم وللتصميم المستخدم في إجراء القياسات. والقرارات لها استخدامان مختلفان هما: قرارات تستخدام للمقارنة بين المفحوصين - قرارات المقارنة-، وقرارات تؤخذ دون إجراء قياس لمفحوصين آخرين -قرارات مطلقة-.

وتميز نظرية إمكانية التعميم بين درسات إمكانية التعميم والقرار. فهدف الأولى هو دراسة إمكانية التعميم والقرار، فهدف الأولى هو دراسة إمكانية تعميم طريقة القياس، في حين أن هدف الأخرى هو التزود ببيانات تستخدم في صنع قرارات تتعلق بالمفحوصين. وتساعد دراسة إمكانية التعميم المصممة جيداً مصمم دراسة القرار في اختيار التصميم الذي يحصل منه على بيانات قابلة للتعميم. ومصمم دراسة القرار الذي يستمد بياناته من دراسة إمكانية تعميم مناسبة سيستخدم نتائج بيانات تحليل التباين لدراسة إمكانية التعميم. وتستخدم هذه النتائج في تقدير معاملات إمكانية التعميم وتباينات الخطأ لتصميمات متنوعة ضمن اعتبارات دراسة القرار. واعتماداً على هذه التقديرات ولاعتبارات تطبيقية يمكن للفاحص أن يختار أكثر التصميمات ملائمة لدراسة القرار.

التمارين:

1/ باستخدام البيانات المدونة في جدول (7-1) وجدول (7-3) أنجز المهام الآتية:

أ- افترض أن الفقرات كانت عينة عشوائية من ملف اختباري كبير. وفي البرنامج الاختباري الفعلي أخذت عينة عشوائية مؤلفة من (20) فقرة وطبقت على المفحوصين. إحسب معامل إمكانية التعميم لهذا الموقف.

ب- افترض أن الطلبة تقدموا للاختبار المؤلف من (20) فقرة المأخوذة من الملف الاختباري المذكور في (أ). وستستخدم نتائج الاختبار في تقدير نسبة الفقرات التي يعرفها التلميذ من الملف. أحسب فترة الثقة المناسبة لمفحوص حصل على الدرجة (20) في الموقف الاختباري الفعلي.

جـ- أحسب معامل إمكانية التعميم المناسب لموقف تقدم فيه الطلبة لاختبار مؤلف من (10) فقرات أخذت من الملف المذكور في (أ).

2/ مختص في القياس مهتم في معرفة ما إذا كان الأطفال الكبار يظهرون تفضيلاً لاختيار أحد بدائل الاستجابة في أسئلة اختيار من متعدد. وقد بنى هذا الباحث أربعة صيغ اختبارية في الحساب تألف كل منها من (20) فقرة، وطبقها جميعاً على (43) مفحوص في الصف الثالث. وكان عدد مرات الإجابة الصحيحة لكل بديل هي نفسها، وصحح الباحث الاختبارات باستخدام المخطط الآتي:

البديل	عدد النقاط
حذف	صفر
أ	1
ب	2
جـ	3
د	4

ويبين الجدول الآتي نتائج تحليل التباين الثنائي:

جدول خلاصة تحليل التباين			
مصدر التباين	درجة الحرية	مجموع المربعات	متوسط المربعات
المفحوصين	42	1915.752	45.613
الصيغ الاختبارية	3	59.791	19.930
الخطأ	126	1782.281	14.145

أحسب معامل إمكانية التعميم الذي يؤشر إلى ما إذا كان التلاميذ يميلون لتفضيل بديل ما وباتساق..

3/ بالاعتماد على مواصفات دراسات D المبينة أدناه، نفذ المهام التالية:

أ- حدد أبعاد الدراسة.

ب- حدد أي الأبعاد - إن وجدت - ثابتة في نطاق التعميم،

جـ- اكتب الصيغة الرمزية للنطاق ولتباين الدرجة الملاحظة.

د- على افتراض أن بيانات دراسة G المتوافرة أجريت باستخدام تصميم دراسة D، بين كيف يمكن حساب عناصر التباين المطلوب.

أ- اختبار نهائي مؤلف من (50) فقرة طبق على فصل دراسات اجتماعية.

ب- صيغة اختبارية قصيرة لاختبار تفهم الموضوع مؤلف من 10 بطاقات طبق على 50 مفحوص، ويتطلب هذا الاختبار من المفحوصين كتابة قصة قصيرة تنطبق على كل بطاقة. وسجلت القصص العشر التي قدمها جوڤنيل ديلينكويت، ودرج كل تسجيل من قبل ثلاثة مختصين نفسيين لدرجة رفض السلطة، وكانت درجة جوڤنيل هي متوسط التقديرات الثلاثة.

جـ- في دراسة للعوامل المرتبطة بالكفاءة الكتابية كتب كل مشترك ثمانية مقالات مقالين لكل نمط، من الأنماط الآتية: وصفي، قصصي، تفسيري، جدلي. واختيرت موضوعات مختلفة لكل نمط، ولكن المشتركين جميعهم كتبوا في الموضوعات نفسها. وصححت الموضوعات لعدد وحدات (i) على أنها مقياس للتعقيد التركيبي. وحسبت الدرجة النهائية للمفحوص بحساب متوسط درجاته على المقالات الثمانية، وصححت المقالات جميعها من قبل مصحح واحد.

د- أجريت دراسة للعوامل المرتبطة بالتحصيل القرائي وطبقت الدراسة على 50 صف دراسي من مستوى الثالث. واستخدم الصف الدراسي على أنه وحدة التحليل، واستخدم متوسط درج الصف كمقياس للتحصيل في اختبار القراءة التحصيلي (ولأهداف إكمال دراسة D افترض أن عدد الأطفال متساوٍ في الصفوف الدراسية جميعها).

4/ أجريت دراسة إمكانية تعميم تهتم بالتفاعل العائلي في المتاحف، إذ تم ملاحظة سلوكات 15 عائلة اختيرت عشوائياً من قبل ملاحظين أثنين لكل عائلة وفي ثلاثة مواقف. وكان أحد المتغيرات التي أهتم بها في الدراسة الوقت الذي يستغرقه الوالدين بدون تفاعل مع الطفل. ويلخص الجدول الآتي خلاصة تحليل التباين على هذا المتغير:

جدول خلاصة تحليل التباين			
مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات
العائلات (P)	2346.955	14	167.639
الملاحظين (I)	3.211	1	3.211
المواقف (J)	1231.622	2	615.811
PxI	18.288	14	1.306
PxJ	2372.377	28	84.726
LxJ	2.688	2	1.344
البواقي	35.311	28	1.261

باستخدام هذه النتائج نفذ المهام الآتية:

أ- خطط الفاحص لإجراء دراسة للارتباطات للفترات التي يستغرقها الوالدين دون التفاعل مع أطفالهم، علماً بأنه تم ملاحظة العائلات من قبل مقدر واحد وفي موقف واحد. أحسب معامل إمكانية التعميم المناسب لهذا التصميم بافتراض أن كل من ملاحظات الفاحص والمواقف متغيرات عشوائية.

ب- إن كان ممكناً من الناحية العملية للفاحص في (أ) زيادة أي عدد من المواقف الملاحظة أو عدد الملاحظين فأيهما تكون زيادته أكثر أهمية.

جـ- افترض أن الفاحص استخدم ثلاثة ملاحظين يمكن لكل منهم ملاحظة العائلات في موقفين على الأكثر. وعلاوة على ذلك فإنه يمكن سؤال كل عائلة من (30) عائلة على الأقل في ستة مواقف. فأي التصميمات يكون قابلاً للتطبيق مع هذه المحددات؟ وأيها ينتج عنه أكبر معامل إمكانية تعميم.

د- افترض أن الفاحص خطط لإجراء الدراسة باستخدام المواقف الثلاثة متغير ثابت. احسب معامل إمكانية التعميم المناسب لهذا التصميم.

هـ- افترض أن الباحث خطط لإجراء الدراسة في متحف مختلف، ولكنه خطط لثلاثة مواقف ومقدرين أثنين كما هو الحال في دراسة D. لأي من الفرعين (د ، هـ) يكون معامل إمكانية التعميم المحسوب في الفرع (د) أكثر ملائمة؟ ولماذا؟

الفصل التاسع

9

معاملات ثبات الاختبارات محكية المرجع

الفصل التاسع

معاملات ثبات الاختبارات محكية المرجع

يتألف الاختبار التحصيلي عادةً من مجموعة فقرات يعتقد بأنها تتطلب المهارات والقدرات التي تمثل أهداف التعليم. وبيّن جليسر (Glaser, 1936) إن درجات الاختبارات التحصيلية تزودنا بنوعين من المعلومات، أحدهما الموقع النسبي لدرجة المفحوص على توزيع الدرجات، وكتابة نتائج الأداء في الاختبار على شكل درجات زائية يوفر مثل هذه المعلومات. والمعيار المستخدم هنا في تفسير الأداء الاختباري يكون من النوع النسبي، والدرجة المُعطاة للمفحوص تسمى القياس معياري المرجع. والنوع الثاني من المعلومات هي درجة اكتساب الطالب لأهداف التعليم، ومثال بسيط لهذا هو درجة النسبة الصحيحة للمفحوص في اختبار يغطي مئة حقيقة في عملية الجمع. فإذا اختيرت فقرات الاختبار عشوائياً من الفقرات الممكنة جميعها، فإن درجة نسبة الإجابات الصحيحة يمكن اعتبارها تقدير لعدد الحقائق التي يعرفها التلميذ وبالتالي يمكن تفسيرها دون معرفة أداء المفحوصين الآخرين على الاختبار، وفي هذا المثال فإن نسبة الإجابات الصحيحة تعد قياس محكي المرجع.

وكما حدد من قبل جليسر فإن التمييز الأساسي بين القياس معياري المرجع والقياس محكي المرجع هو المعيار الذي ينسب إليه الأداء. وظهر مصطلح القياس محكي المرجع في تعريف جليسر على أنه «.. هو تقييم للسلوك الطرفي أو المحكي». ويستخدم المصطلح قياس محكي المرجع كبديل للمصطلح الأكثر إزعاجاً «قياس السلوك محكي المرجع»، والذي يشير إلى أنه يمكن تفسير القياس من خلال سلوكيات محكية يمكن أن يظهرها الطالب أو المفحوص. والمصطلح معياري في القياس معياري المرجع يشير إلى حقيقة أن هذه القياسات تفسر من خلال معايير الاختبار (أنظر الفصل 19- مناقشة معايير الاختبار.

وفي عام 1969 أكد كل من بوبام وهيوسك **(Popham & Husek, 1969)** أن التقنيات المتوافرة في بناء الاختبار، وتقدير الثبات، وتقييم الصدق، وتحليل الفقرات تم تطويرها بشكل موسع للاختبارات معيارية المرجع. وتتطلب الاختبارات محكية المرجع من وجهة نظرهم تقنيات جديدة في كل من هذه المجالات. وقد صاغ ليفنجستون **(Livingston, 1972)** وهامبلتون ونوفيك **(Hambelton & Novick, 1973)** معاملات جديدة لثبات القياس المحكي. ومنذ نشر هاتين الدراستين

كُرسٌ عدد كبير جداً من أدب القياس لصياغة هذه المعاملات. وتم مراجعة أدبيات الموضوع في هذا الفصل، ومن المراجع الأخرى للموضوع : (Berk, 1980, a,b) وبرينان (Brenan, 1980) وسبكوفياك (Subkoviak, 1980) وتراوب ورولي (Traub & Rowley, 1980).

استخدامات القياس محكي المرجع:

عندما يتم تفسير القياس على أنه محكي المرجع من خلال السلوكات المحكية التي يظهرها المفحوص فإن الاختبار المحكي يجب تصميمه ليسمح بمثل هذا النوع من التفسير، والعديد من كتبة الاختبارات يبدأ بناء الاختبار محكي المرجع بتحديد المجال أو مجتمع المهام التي يجب أن يكون المفحوص قادراً على أدائها (أنظر1974، Millman لوصف طرائق تحديد المجال). ويبني الاختبار بمعاينة المهام من مجال المهام، وأثنان أو أكثر من هذه الاختبارات تسمى اختبارات متكافئة أو متوازية أسمياً أو عشوائياً. واشتقت هذه الأسماء من حقيقة أن مثل هذه الاختبارات تعد متوازية في المحتوى ولكنها ليست متوازية تماماً، فليست بحاجة لتساوي متوسطاتها وتبايناتها أو ارتباطاتها مع الدرجات الشاملة. وهنالك هدفين مرتبطين بهذه الطريقة في بناء الاختبار محكي المرجع على نتائج الاختبار (Hambelton, etal, 1978a). الأول: تقدير درجة النطاق، وبدقة أكثر نقول أن درجة النطاق للمفحوص P هي نسبة الفقرات في النطاق التي يمكن أن يجيب عنها المفحوص P إجابة صحيحة[1]. ودرجة النطاق هذه هي الدرجة الشاملة في نظرية إمكانية التعميم. وهنا يتألف نطاق التعميم من جميع فقرات النطاق الكلي للفقرات، ودرجة النطاق (أو الدرجة الشاملة) للمفحوص P هي متوسط درجاته على جميع الفقرات. والثاني: في موقف السيطرة، والذي يقسم فيه متصل درجة النطاق إلى فئات سيطرة تحدد بدرجات قطع عددها (k-1)، وتستخدم نتائج الاختبار الملاحظة في تصنيف المفحوصين إلى فئات السيطرة

والمثال الأكثر شيوعاً هو درجة قطع واحدة وفئتين مسيطر وغير مسيطر. وبما أن هنالك هدفين مختلفين تم تحديدهما لدرجات الاختبار محكي المرجع (تقدير درجة النطاق وتصنيف السيطرة) فإنه يتطلب الحاجة إلى وجود طريقتين مختلفتين لحساب ثبات كل منهما (Traub & Rowley, 1980).

نظرية الثبات لدرجة النطاق المحسوبة:

خذ كمثال موقف تقويم منهج إسباني جديد يريد المقوِّم بوساطته تقييم المدى اللفظي

---

(1) سنهتم بهذا الفصل في موقف القياس التي تكون درجة النطاق فيها متغير متصل. والقراء المهتمين بالمواقف الأخرى التي للمفحوصين قدرة على الإجابة عن جميع الفقرات أو المهام أو لا شيء منها يمكنهم الرجوع إلى تراوب وروليي (Traub & Rowely, 1980).

المتوافر لدى كل مفحوص. والنطاق هنا عبارة عن الكلمات الإسبانية المذكورة في المنهج جميعها. ولأسباب تطبيقية واضحة يمكن اختبار التلاميذ بعدد محدود من الكلمات فقط، وينوي المقوِّم استخدام درجة النسبة الصحيحة الملاحظة على أنها تقدير لدرجة النطاق للتلميذ: أي نسبة الكلمات التي يعرفها التلميذ. ولكن ما درجة دقة التقريب لنسبة الاجابات الصحيحة بالنسبة لدرجات النطاق؟ فإن كان اهتمامنا بالاختبارات المبنية بالمعاينة العشوائية من عينة لا نهائية من الفقرات فإن الإجابة على السؤال تكون من خلال نظرية إمكانية التعميم المبينة في الفصل 8 للتصميمات أحادية البعد. وعلاوة على أنه يمكن تعديل هذه النتائج لتتوافق مع ملف لا نهائي من الفقرات (Sirotnik, 1972) والمعاينة الطبقية (Rajaratnam, Cronbach & Gleser, 1965).

ومن وجهة نظر نظرية إمكانية التعميم فعندما يطبق مقوِّم الاختبار اختبار الألفاظ الإسبانية فإنه تجري دراسة P يتقاطع فيها المفحوصين مع الفقرات. وتم الترميز لهذا التصميم بـ Pxi وأعطي مسمى تصميم 2 للتصميمات أحادية البعد. وهنا يستخدم المقوِّم درجة النسبة الصحيحة الملاحظة على أنها تقدير لدرجة نطاق المفحوص P (أو الدرجة الشاملة). وفي التحقق من مدى جودة تمثيل الدرجة الملاحظة لدرجة النطاق فإن هنالك كميتان يجب الاهتمام بهما:

الأولى: تباين الخطأ (أو جذره التربيعي والذي هو الخطأ المعياري للقياس)، والثانية هي معامل إمكانية التعميم. ولأسباب ستناقش في الحال يفضل استخدام تباين الخطأ، ولذلك دعنا نبدأ بمناقشة تباين الخطأ عندما تكون درجة نسبة الإجابة الصحيحة الملاحظة تقدير لدرجة النطاق. ويعد معالجة درجة نسبة الإجابة الصحيحة على أنها تقدير لدرجة النطاق هو مثال لقرار مطلق، ويجب أن يسترجع القارىء من فصل (8) إن القرار المطلق هو قرار يعتمد على درجة المفحوص دون الرجوع إلى درجات المفحوصين الآخرين.

وكمراجعة للمفاهيم المناسبة لهذا النقاش فقد يرغب القارىء فحص البيانات المقدمة في جدول (7-1) التي تحتوي استجابات 10 تلاميذ ($n_p = 10$) على اختبار مؤلف من 6 فقرات ($n_i = 6$)، ويبين جدول (7-2) حساب معامل ثبات كيودر- ريتشاردسون 20 لهذه الفقرات. وجدول (7-3) يبين كيفية حساب هذا المعامل وذلك بطريقة تحليل التباين لهويت. ويتضمن جدول (9-1) ملخصاً لهذه المعلومات والمهمة في توضيح كيفية الحصول على تقدير الثبات فيما لو تم تفسير الدرجات على أنها محكية المرجع. وتحديداً فإن تحليل تباين الأفراد x الفقرات يؤدي إلى ثلاث كميات هي:

مربع متوسطات الأفراد (MS_p)، ومربع متوسطات الفقرات (MS_i)، ومربع متوسطات

تباين البواقي (MSr). وفي فصل 8 تم تجميع مربع المتوسطات المحسوبة للحصول على تقديرات لعناصر معاملات إمكانية التعميم. وكما هو ملاحظ في دراسة D أحادية البعد (وكما هو الحال في مثالنا) وبمعالجة الفقرات على أنها البعد، فإن عناصر التباين للأفراد والفقرات

جدول (9-1): خلاصة استجابات الفقرة ونتائج تحليل التباين من الجداول (7-1) إلى (7-3) للفقرات الستة والمفحوصين العشرة

المفحوصين	1	2	3	4	5	6	المجموع
	الفقرات						
1	0	0	0	0	0	0	0
2	0	0	0	0	1	0	1
3	1	0	1	1	1	0	4
4	1	1	1	1	1	1	6
5	1	1	1	1	1	1	6
6	0	0	1	0	0	0	1
7	0	0	1	1	1	0	3
8	0	0	0	1	0	0	1
9	1	0	1	1	1	0	4
10	0	1	0	1	0	1	3
p	.40	.30	.60	.70	.60	.30	

SV	MS	df	S.S
الأفراد	0.7574	9	6.817
الفقرات	.2967	5	1.483
البواقي	.1485	45	6.683

يمكن حسابها بـ:

$$\sigma^2_p = \frac{EM_r - EM_p}{n_i}$$

$$\sigma_i^2 = \frac{MS_r - MS_i}{n_p} \quad \text{و} \quad MS_r = \sigma_i^2$$

وفي مثالنا فإن σ_i^2 هو تباين صعوبات الفقرة لنطاق الفقرات، و σ_e^2 يعكس خطأ القياس وحقيقة أن أرجحية الإجابة الصحيحة لمفحوص عن فقرة إجابة صحيحة ستتغير من فقرة لأخرى، وباستخدام المعلومات المبينة في جدول (9-1) لهذا المثال فإن:

$$\sigma^2_p = \frac{0.1485 - 0.7574}{6}$$

$$= 0.1015$$

$$\sigma^2_i = \frac{0.1485 - 0.2967}{10}$$

$$= 0.0148$$

وتباين الخطأ المناسب لهذا التصميم هو $\sigma_i^2 + \sigma_e^2/n_i$ مع أن استخدام الصيغة الآتية أسهل في حساب تباين الخطأ:

$$\hat{\sigma}_e^2 + \hat{\sigma}_i^2 = \frac{(1-X_{pI})\, x_{pI} \bar{X}_p}{(1-n_i)\, n_p} \quad (1\text{-}9)..........$$

حيث ترمز n_p إلى عدد المفحوصين و x_{pI} إلى نسبة الإجابات الصحيحة للمفحوص P.

والقيمة المحسوبة لتباين الخطأ يجب أن تقع في الفقرة من صفر إلى $\frac{0.25}{1-n_i}$. لذلك فإن تباين الخطأ لا يمكن أن يكون أكبر من 025.، ومن ثم فإن الخطأ المعياري يجب أن يكون أقل من $\sqrt{0.25}$ أو 0.158، رباستخدام المعادلة (9-1) وبيانات جدول (9-1) نحصل على $\sigma_e^2 + \sigma_i^2 =$

$$\frac{0.25 + 2211 + 0.1411 + 0.25 + 0.1411 + 0.00 + 0.00 + 0.2211 + 0.1411 + .00}{(5)\,(10)}$$

$$= 0.0273$$

وحيث أن تباين الخطأ هو $n/\ \sigma_e^2 + \sigma_i^2$ فإن هنالك معامل إمكانية تعميم واحد ممكن لدرجة النطاق هو:

$$\rho_L^2 = \frac{\sigma^2_p}{n_i/\,(\sigma_e^2 + \sigma_i^2) + \sigma_p^2} \quad (2\text{-}9)..........$$

ولمثالنا فإنه يساوي:

$$= \frac{0.1015}{0.0273+0.1015} = 0.79$$

وهذا المعامل يساوي 1 إذا كانت درجة الإجابة الصحيحة الملاحظة لكل مفحوص تساوي درجة النطاق للمفحوص. لذا فإن المعامل الكبير يؤشر إلى تقدير دقيق لدرجات النطاق. لنرى هذا، افترض أن المفحوصين جميعهم حصلوا على درجة النطاق نفسها، أي أن $\sigma^2 p$ = صفر، وبالتالي فإن ρ_L^2 = صفر. والآن إذا تألف الاختبار من 11 فقرة فإن أقصى قيمة لتباين الخطأ = 0.025، وإذا كان تباين الخطأ المحسوب = .001 فقط، فإن الدرجات الملاحظة تكون تقديراً دقيقاً وعادلاً لدرجة النطاق المفردة. لذلك إن كان تباين الخطأ المحسوب = .025، فإن الدرجات الملاحظة لا تكون تقديراً دقيقاً جداً لدرجة النطاق.

ومع ذلك ففي كلا الحالتين ρ_L^2 = صفر، لذا فمن الواضح ان معامل امكانية التعميم المنخفض لا يشير دائماً الى قياس غير دقيق. ووجهة نظرنا ان تباين الخطأ المحسوب باستخدام نظرية امكانية التعميم نفسيره اسهل من تفسير معامل امكانية التعميم على انه مقياس لدرجة النطاق المحسوب في الاختبارات محكية المرجع.

والتفسير معياري المرجع لدرجة النطاق في بعض الحالات هو المهم . وفي هذه الحالة فان سؤال الثبات يدور حول درجة جودة تنبؤ درجة الاجابة الصحيحة الملاحظة (X_{pl}) بدرجة نسبة الاجابة الصحيحة للنطاق (μ_p). فان كانت x_{pl} تتنبأ بـ μ_p بالضبط فان اي تفسير معياري لـ x_{pl} سيكون مكافئاً للتفسير المناظر لـμ_p . ومعامل امكانية التعميم يكون:

$$\frac{\sigma^2 j}{n_1/\sigma_e^2 + \sigma_p^2} = \rho_L^2 \qquad (3-9)\ldots\ldots\ldots\ldots\ldots\ldots\ L$$

وسيساوي واحد عندما يكون التنبؤ تاماً، وصفر عندما لا تفيد X_{pl} في التنبؤ عن μ_p. ويكون معامل امكانية التعميم مفيداً في التفسير معياري المرجع للاختبارات محكية المرجع. والتقدير في مثالنا هو:

$$0.80 = \frac{0.1015}{6/0.1485 + 0.1015} = \rho^2{}_{1*}$$

وكما لوحظ في الفصل الثامن للبيانات الثنائية فان $\rho_{L}^{2}(x_{pl}, \mu_p)$ تكافيء مربع الارتباط بين درجة النطاق ودرجة النطاق المحسوبة. وكما لاحظنا في الفصل 8 فان حساب هذا العامل يتطلب بناء ثلاثة صيغ اختبارية وتطبيقها.

نظرية الثبات لتصنيف السيطرة:

ان دقة درجات الاختبار كتقدير لدرجات النطاق تكون اقل اهمية عند استخدام الاختبار في تصنيفات السيطرة. لنعد الى اختبار الالفاظ الاسبانية ثانية، وافترض ان نقطة القطع في هذا الاختبار هو ان يحصل المفحوص على درجة القطع او درجة اعلى منها ليصنف كمؤهل لدراسة المحادثة الاسبانية المتقدم، والذين حصلوا على درجة اقل من درجة القطع عليهم اعادة مساق المستوى التمهيدي ليتسنى لهم الانتقال الى المستوى التالي. وفي هذه الحالة لا يهتم الفاحص بدقة درجات النطاق المحسوبة اكثر من توافقها مع أو دقة قرارات التصنيف التي اجراها المقدرون: لترى هذا، افترض ان جميع المفحوصين حصلوا على درجات نطاق تجاوزت درجة القطع للسيطرة وصنف الجميع على انهم مسيطرون ، وذلك من خلال بيانات الاختبار. وحيث ان درجة النطاق ودرجة التصنيف الملاحظة متوافقين فان دقة الدرجات الملاحظة كتقدير لدرجات النطاق قد تكون غيرملائمة.

اتساق القرار المقدر باختبارين او موقفين اختبارين:

يهتم اتساق القرار بالمدى الذي تعطي فيه قرارات متشابهة من خلال مجموعتين مختلفتين من القياس. ويمكن ان يشير الاتساق الى تطابق القرارات اعتماداً على صيغتين للاختبار او تطبيقين مختلفين للاختبار نفسه. وبشكل أولي سنرى في هذا الفصل توضيح عند استخدام صيغتين اختباريتين. على سبيل المثال: تطبيق صيغتين مختلفتين على 100 مفحوص قد يؤدي الى النتائج المبينة في شكل (9-1). فعندما تكون نتائج كلا الصيغتين تؤشران بأن المفحوص صنف مسيطراً، فان القرارات لهذا المفحوص تكون متوافقة. والاحتمال المحسوب لاتساق قرار السيطرة هو عدد المفحوصين الذين صنفوا باتساق على انهم مسيطرين مقسوماً على عدد المفحوصين الكلي. وفي مثالنا كان اتساق قرار السيطرة لـ 20 مفحوص، لذا فان الاحتمال المحسوب لاتساق تصنيف السيطرة يكون $P_{11} = 0.20$ وعندما تكون نتائج كل صيغة تؤشر تصنيف غير مسيطر للمفحوص فان القرارات تكون متسقة ايضا ومن المئة طالب كان اربعون فقط تصنيفهم باتساق على انهم غير مسيطرين، لذا فان الاحتمالية المحسوبة لاتساق غير مسيطر Poo=0.4 والاحتمالية للاتساق الكلي هو: $P = P_{11} + P_{oo}$

وفي مثالنا P=.2+.4=0.6 وبالتالي فإن 60% من المفحوصين كان تصنيفهم متسقاً.

وفقد اقترح هامبلتون ونوفيك (1973) $\hat{P}$ على انها مقياس لاتساق القرار. والتصنيفات

القرار المعتمد على الصيغة 1

القرار المعتمد على الصيغة 2	غيرمسيطر $0.1 = P_{01}$	غيرمسيطر $0.4 = P_{00}$	غيرمسيطر $0.5 = P_{01}$
	$0.2 = P_{11}$	$0.3 = P_{10}$	مسيطر $0.5 = P_{10}$
	$0.3 = P_{01}$	$0.7 = P_{10}$	

شكل (9-1):احتمالات اتساق التصنيف لصيغتين اختباريتين

الاخرى جميعها سواء أكانت (مسيطر– غير مسيطر) أو (غير مسيطر – مسيطر) هي قرارات غير متسقة والاحتمالية المقدرة للتصنيف غير المتوافق هي:

$$0.4 = P - 1$$

العوامل المؤثرة في اتساق القرارات:

هنالك اربعة عوامل تؤثر في اتساق القرار:

1/ طول الاختبار.

2/ موقع درجة القطع على توزيع الدرجات.

3/ تعميم درجة الاختبار.

4/ تشابه التوزيع لدرجات صيغتي الاختبار.

لتوضيح تأثير هذه العوامل على احتمالية اتساق القرار سنعرض نتائج بيانات افتراضية لصيغتي اختبار متبادلتين " درجات الاختبار للصيغ المتبادلة لها التوزيع نفسه والعلاقة بدرجة المجال (تعد التبادلية صيغة اقوي للتكافؤ من التوازي التام. فالصيغ المتبادلة هي متوازية تماماً، ولكن المتوازية تماماً ليس من الضروري ان تكون متبادلة). وعلى الرغم من ان نتائج جدول (9-2) تصنف صيغتي اختبار متبادلتين الا ان الاتجاه العام يطبق على الصيغ المتكافئة العشوائية ويبين الجدول (9-2) قيمة (أ): احتمالية اتساق القرار لصيغتي اختبار

مؤلف من خمس فقرات. واخر مؤلف من (10) فقرات، مع تدوين قيم P لعدة درجات ممكنة لكل طول اختباري . ويوضح جدول (9-2) ايضاً انه ولزوج الصيغ الاختبارية يميل قرار الاتساق لان يكون اقل عندما تكون درجة القطع اقرب الى مركز توزيع درجات الاختبار، وفي جدول -2) (9 كان متوسط النسبة المئوية للاجابة الصحيحة= 0.4 لكلا الصيغتين ، والأقل في اتساق القرار كان لدرجة قطع = 0.4 . ويوضح جدول (9-2): أثر التغير في درجة القطع وطول الاختبار على ثبات P (اتساق القرار): ويبين الجدول ايضا ان امكانية التعميم تزداد بزيادة

عدد فقرات الاختبار	درجة القطع (مقياس النسبة المئوية للاجابة الصحيحة)(a)				
	P2(b)	0.2	0.4	0.6	0.8
5	0.40	0.81	0.66	0.68	0.81
10	0.57	0.83	0.71	0.77	0.90

(a) : متوسط درجة النسبة المئوية للاجابة الصحيحة= 0.40 لكلا الاختبارين

(b) المعامل $\rho 2$ هو مربع الارتباط بين الدرجة الملاحظة ودرجة النطاق لكلا الصيغتين، مقتبس عن هونيه .H) Hunyh) . الحساب والاستدلال لكلا معاملي الثبات في اختبار السيطرة كان بالاعتماد على نموذج ثنائي الجد - بيتا .

طول الاختبار، وهذه تؤثر في زيادة اتساق القرار وتؤدي زيادة طول الاختبار إلى زيادة في اتساق القرار ودرجات لقيم درجات القطع الاربع جميعها، الا ان الزيادة ليست نفسها لكل درجات القطع، واتجاه الزيادة الاكبر يكون لدرجات القطع الاكبر ولكنها لا تعمم لمواقف اخرى. ويمكننا استنتاج ان مدى تأثير طول الاختبار على اتساق القرار يعتمد على الموقف نفسه فقط.

والنتائج المبنية في جدول (9-2) لكل طول اختبار، اذ تشير كل منρ^2 ودرجات القطع الى مجموعة المفحوصين نفسها. ويبين جدول (9-3) قيم P لمجموعات مفحوصين مختلفة، فعلى سبيل المثال يحتوي السطر الاول على قيم P لمجموعة مؤلفة من خمسة مفحوصين، وكل مجموعة لها متوسط قيمته 3.0 لكل صيغة، ولكن تباين درجات النطاق وبالتالي امكانية تعميم درجات الاختبار على درجات النطاق يتغير عبر المجموعات. والاتجاه العام في جدول (9-3) مشابه للاتجاه في جدول (9-2)، فازدياد امكانية التعميم يميل لزيادة اتساق القرار لذا فان صيغتين للاختبار تميل لانتاج قرارات اكثر اتساقا لمجموعة تتميز بدرجات نطاق غير متجانسة اكثر مما هو لمجموعة درجات نطاقها متجانسة.

p21 (b)					عدد فقرات الاختبار(a)
0.9	0.7	0.5	0.3	0.1	
0.90 0.94	0.78 0.91	0.69 0.91	0.63 0.93	0.57 0.96	3.0 4.8

(a) : درجة القطع معبر عنها على متصل الدرجات الكلي = 3 لمدخلات الجدول جميعها

(b) : يشير المعامل ρ^2 الى مربع الارتباط بين الدرجات الملاحظة ودرجات النطاق لكلا الصيغتين . مقتبس عن H. Hunyh . Computation and inference for two reliability indices in mastery testing based on the beta- binomial model, in Hungh & J.C. saunders. Solutions for some technical problems in domain - referenced mastery testing (Final Report, Project NIE.G. 78- 0087). National institute of Education department of Health Education, and welfare. Adapted by permission.

وهذا مبين بشكل واضح في السطر الأول اذ يزداد اتساق القرار من اليمين الى اليسار بازدياد تباين درجات نطاق المفحوصين. ايضا من المحتمل لاتساق القرار أن يكون اقل لمجموعة تتميز بامكانية تعميم اكبر لدرجاتها، وهذا مبين في السطر الثاني من جدول (3-9). ويبين هذا الجدول ايضا أن P تميل لان تكون اقل لمجموعة متوسط درجاتها اقرب الى درجة القطع وكما هو مبين بنتائج كل عمود، ففي كل عمود كان اتساق القرار اقل للمجموعة المذكورة بالسطر الاول اكثر مما هو للمجموعة في السطر الثاني. وحسب هذه النتائج بافتراض درجة قطع 3 نقاط على تدريج الدرجة الخام. فكل مجموعة في السطر الاول لها متوسط درجات = 3 ، في حين كل مجموعة في السطر الثاني كان لها متوسط درجات = 4.8. وضمن العمود كان اتساق القرار اقل للمجموعة في السطر الاول إذ أن متوسطها كان اقرب الى درجة القطع. ويبين جدول (3-9) ايضا ان اتساق قرار اساسي يمكن ان يظهر حتى عندما يكون امكانية تعميم الدرجات منخفض، وهذا مبين في العمود الاول من الجدول. فللمجموعة ذات متوسط 4.8 كان ما نسبته 97.7 من القرارات متسقا حتى عندما كانت امكانية تعميم الدرجات منخفضة نوعاً ما، وظهر هذا لان معظم الدرجات كانت أعلى من درجة القطع على كل صيغة، لذلك فان معظم القرارات متساوية مع قرار السيطرة. ويمكن تبيان انه اذا كانت الدرجات متبادلة، فان اقل قيمة محتملة لـ $p = 0.5$، وهذه القيمة تظهر عندما تكون درجة القطع تقع على وسيط التوزيع والارتباط بين الصيغتين المتبادلتين = صفر وعامل اخر يؤثر في اتساق القرار هو التشابه بين توزيعات درجات اختبار كلا الصيغتين، وعندما

تتساوى الاشياء الاخرى فان اتساق القرار يميل لان يكون اقل عندما لا تتشابه توزيعات درجات الاختبار.

وقد اجريت اقتراحات عديدة في الدراسات لتحويل P الى قياس اكثر قابلية للتفسير لاتساق القرار. اثنان من هذه التحويلات مبينة في هذا الجزء، وبدلا من محاولة تفضيل احدى التحويلات على الاخرى نقترح انهما طريقتان بسيطتان ومختلفتان في تحويل P ، لذا فان المعاملات الجديدة يعبر عنها على تدريجات عليها بعض النقاط.

وقد اقترح كل من سوا ميناثان وهامبلتون والجينا (Swaminathan; Humbelton & Algina, 1974) استخدام معامل كابا لكوهين:

$$k = \frac{P_c - P}{P_c - 1} \quad \text{.....................(9- 4)}$$

حيث ترمز P_c الى الفرصة المحتملة لاتساق القرار، وهو ان الاحتمالية للموقف الافتراضي الذي تكون فيه الدرجات على الصيغتين مستقلة احصائيا، ويشار اليه احياناً على انه اتساق الصدفة. وبحسب استخدام الصيغة $P_{00} \times P_{00} + P_{01} + P_{10} = P_c$ ، ففي هذا التعبير P_{10} تمثل احتمالية تصنيف مسيطر على احدى الصيغ و P_{01} تمثل احتمالية تصنيف السيطرة على الصيغة الاخرى وكذلك تصنيف P_{00} تمثل احتمالية غير مسيطر على كلا الصيغتين . ولبيانات جدول (9-1) فان $0.5 = P_{10}$, $0.3 = P_{01}$, $0.5 = P_{00}$ ، و 0.7 , $0.5 = 0.7 \times 0.5 + 0.3 \times 0.5 = Pc$ P_{00} = وهذا يعني ان 50% من القرارات تكون متسقة حتى عندما تكون درجات صيغتي الاختبار مستقلة احصائيا ويمكن تصور اتساق الصدفة على انه خط اساسي للحكم على الكم الفعلي من الاتساق الملاحظ على كلا الصيغتين، والمقام في K و 1-Pc يمثل اقصى زيادة ممكنة في اتساق القرار اعلي وأدنى من اتساق الصدفة. ويمثل البسط P - Pc الزيادة الفعلية عن اتساق الصدفة. لذا فانه يمكن تفسير k على انها الزيادة في اتساق القرار الذي توفره القرارات اكثر من الصدفة المعبر عنها على انها الزيادة الممكنة في اتساق القرار على اتساق الصدفة، ويكون المعامل k= صفر عندما لا يكون هنالك زيادة، و1.0 عندما يكون هناك اقصى زيادة ممكنة. ولبيانات شكل (9-1) فان $K = \frac{0.5 - 0.6}{0.5 - 1} = 0.2$ ، اي بزيادة 20% من الزيادة الكلية المحتملة اكثر من اتساق الصدفة تم ملاحظتها للقرارات المعتمدة على كلا الصيغتين.

ويجب ملاحظة ان K= صفر لا تعني ان القرارات غير متسقة ولا تستحق الاهتمام. ويمكن تفسير K= صفر على انها لا اتساق اكثر مما يعتمد على درجات اختبار مستقلة احصائيا، وقد يكون هذا الاتساق اساسي، تذكر انه لصيغتي الاختبار المتبادلتين يكون على الاقل 50% من القرارات متسقة. كذلك فان 1-k يمكن تفسيرها لتعني ان القرارات متسقة كتلك المعتمدة على درجات غير مستقلة احصائيا. ويمكن افتراض قيم سالبة لـK لتناظر الموقف الذي تكون فيه علاقة عكسية بين درجات كلا الصيغتين.

في البداية اشرنا الى انه وعندما تكون الصيغتين متبادلتين فان اقل احتمالية لا تساق القرار تكون 0.5، وهذه القيمة الاقل ستظهر فيما لو كانت درجات الصيغتين مستقلة احصائيا وكانت درجة القطع تقع على وسيط التوزيع الخاص بالصيغتين واعتمادا على هذه القيمة الاقل، نستخدم P^* كبديل لـ K:

$$P^* = \frac{0.5 - P}{0.5 - 1} = 1 - 2P \qquad (5 - 9)$$

ومثل K ، يمكن تفسير P^* لتعني ان القرارات متسقة مثل تلك التي تعتمد على اختبارات غير مستقلة احصائياً تماماً. وتساوي P^* صفر عندما تكون القرارات ليست اكثر اتساقا مما لو كانت درجات الاختبارات مستقلة احصائياً من تلك التي لها التوزيع نفسه لكل صيغة ودرجة قطع مساوية لوسيط توزيع الدرجات. ويجب ملاحظة ان P^* قد تكون سالبة عندما يكون توزيع درجات الاختبار الحقيقية غير متشابه تماماً . ولبيانات شكل (9-1) فان $P^* = 2 X (0.6) — 1 = 0.2$ مشيراً الى تحسن (زيادة) في اتساق القرار بنسبة 20% عما هو في حالة الصيغ الاختبارية المتبادلة المستقلة احصائيا، وبدرجة قطع تساوي وسيط التوزيع.

العوامل المؤثرة في K و P^*:

مثلP: احتمالية اتساق القرار، فان كل من P^* و K تتأثران بامكانية تعميم درجات الاختبار، وموقع نقطة القطع على توزيع الدرجات وتشابه التوزيعات والجدولين (9-4) و (9-5) يبينان قيم P^* و K للمواقف نفسها المستخدمة في الجدولين (9-2) و (9-3) . فالجدول (9-4) يبين الاثر على K وP لتغير طول الاختبار على إمكانية تعميم درجات الاختبار وتغير درجة القطع. وهذه النتائج هي للمفحوصين أنفسهم. والجدول (9-5) يبين الاثر على K وP^*. لتغير تركيب المجموعة على امكانية تعميم المدرجات وموقع درجة القطع في توزيع درجات الاختبار. ونخلص بالقول الى ان P^* تتأثر تماما بالطريقة نفسها التي تتأثر بها P. ويميل المعامل P^* لان يكون اقل عندما تكون درجة القطع قريبة من متوسط التوزيع. وعلى الاغلب فانP^* تزداد بازدياد امكانية التعميم لدرجات الاختبار ويميل

جدول (9-4) : اثر تغير درجة القطع والتغير في طول الاختبار على ثبات P^* و K [(a)]:

المعامل	عدد الفقرات	$\rho^{2(b)}$	0.2	0.4	0.6	0.8
		درجة القطع (مقياس النسبة الصحيحة)				
P*	5	0.40	0.62	0.32	0.36	0.62
	10	0.57	0.66	0.42	0.54	0.80
k	5	0.40	0.20	0.28	0.29	0.23
	10	0.57	0.31	0.41	0.39	0.28

(a) : متوسط درجة الاجابة الصحيحة = 0.4 وللاختبارات جميعها

(b) : المعامل ρ^2 هو مربع الارتباط بين الدرجة الملاحظة ودرجة النطاق لكلا الصيغتين (مقتبس عن نفس المرجع السابق).

المعامل K لان يكون اكبر عندما تكون درجة القطع قريبة من مركز التوزيع وتزداد ايضا بازدياد امكانية تعميم درجات الاختبار. وفي ادبيات الموضوع ستجد نقاشاً حول اي المعاملين هو الانسب بدرجة اكبر. ووجهة نظرنا هي ان كلاهما طرائق بديلة لتفسير P، وعند فهم طبيعة التفسير فلا نفضل احداهما على الاخرى.

جدول (9-5) : تاثير تغيير تركيب المجموعة على ثبات درجاتP^* و$K^{(a)}$ ومتوسطاتها:

المعامل	متوسط الاختبار	0.1	0.3	0.5	0.7	0.9
		$\rho^{2(b)}$				
P*	3.0	0.14	0.26	0.38	0.78	0.8
	4.8	0.92	0.86	0.82	0.82	0.88
k	3.0	0.07	0.21	0.36	0.54	0.79
	4.8	0.02	0.12	0.27	0.49	0.78

(a) : درجة القطع معبر عنها على تدرج الدرجة الكلية = 0.3 لمدخلات الجدول جميعها،

(b) : المعامل ρ^2 يشير الى مربع الارتباط بين الدرجات الملاحظة ودرجات النطاق لكلا الصيغتين (مقتبس عن نفس المرجع السابق)

والعامل الاخير المؤثر على P^* و K هو تشابه توزيع الدرجات على صيغتي الاختبار،

فكلاهما يميل للانحدار كلما بعد التشابه بين التوزيعين، فضلا عن ان K أو *P لا تصلا الحد العلوي الا اذا كانت احتمالات السيطرة وعدم سيطرة الهامشية هي نفسها لكلا الصيغتين.

معاملات ليفنجستون وبرينات - كين :-

ليس بعيداً اعتبرنا معالم اتساق القرار الذي يعالج تصنيفات غير المسيطر متساوية الجدية، افترض انه وعند درجة القطع 0.7 صنف المفحوص A ودون اتساق باستخدام نسبة الاجابات الصحيحة وكانت 0.65 على احدى الصيغ و 0.75 على الاخرى. وصنف المفحوص B دون اتساق بحصوله على درجات 0.20 و 0.90 ، وعندما اراد مطور الاختبار ان يعكس مقدار فجوة التصنيف في الحكم على ثبات القرارات فان هنالك معاملين يفيدان في الحكم طورهما كل من ليفنجستون(Livingston, 1972) وبرينان وكين (Brenan & Kane, 1977) . وقد قدم كل من برينان وكين معالجة اكثر أهمية (اساسية) لتطوير معاملات الاتساق هذه اكثر مما سنوفره هنا. تذكر ان توزيع الدرجات على الاختبارين متشابهة، ويعتمد اتساق القرار علي عاملين: امكانية تعميم درجات الاختبار وموقع درجة القطع على توزيع درجات الاختبار . علاوة على ذلك ومن الحدس يزداد اتساق القرار كلما يزداد كل من امكانية تعميم درجات الاختبار و / أو الفروق (المسافة) بين متوسطات درجات الاختبار ودرجات القطع. وتقاس هذه الخاصية بشكل مناسب ومحدد بوساطة معامل ليفنجستون ، وكما يأتي:

$$\frac{^2(n_i c - \mu_T) + \sigma_T^2}{^2(n_i c - \mu_X) + \sigma_p^2} = k\ ^2(X,T) \qquad (6 - 9)..............$$

وتستخدم الرموز X و T منذ ان طور ليفنجستون هذا المعامل في سياق النظرية التقليدية/ ويشير الرمز C الى درجة القطع المعبر عنها على التدريج المئوي للدرجات، و ni هي عدد فقرات الاختبار. وحساب $K^2(X\ ,\ T)$ سهل جداً، فإن توافر قياسين لكل مفحوص تستخدم الصيغة:

$$\frac{(n_i c - \hat{\mu}_{X'})(n_i c - \hat{\mu}_X) + \hat{\sigma}_{X'} + \hat{\sigma}_X\ \hat{\rho}_{XX'}}{[^2(n_\iota - \hat{\mu}_{X'})_+ \hat{\sigma}^2_{X'}]\ \ [^2(n_\iota c - \hat{\mu}_X) + n_2/\hat{\sigma}^2_{X'}]} = k\ ^2(X,T) \qquad (7 - 9)...............$$

وهنا يمثل $\rho_{XX'}$ الارتباط بين درجات الصيغتين و μ تمثل متوسط الدرجات الخام، وان توافرت صيغة اختبارية واحدة تستخدم المعادلة:

$$\frac{^2(n_i c - \mu_X) + (KR20)\ \sigma^2 x}{^2(n_i c - \mu) + \sigma^2 x} = k\ ^2(X,T) \qquad (8 - 9)...............$$

كمثال في بيانات جدول (9-1) وعند درجة قطع = 0.67 ، فإن:

$$0.85 = \frac{^2[0.67 \times (6) - 2.9] + (0.80)\, 4.08}{^2(0.67)(6) - 2.96 + 4.08} = \hat{k}\,.\,^2(X,T)$$

ان العامل الذي يؤثر على اتساق القرار هو تشابه توزيع درجات كلا الصيغتين. فعندما يتم بناء الصيغتين من خلال معاينة نطاق الفقرات ينتج توزيعات مختلفة، وذلك لان الفقرات مختلفة في صعوباتها وتؤدي الى صيغ مختلفة، وقد اقترح برينان وكين (1977) معامل اتساق القرار الذي يأخذ بعين الاعتبار الاختلاف في صعوبات الفقرات.

$$M(C) = \frac{^2(c - \mu) + \sigma^2 p}{n_i/(\sigma^2 e + \sigma^2{}_i) + {}^2(c - \mu) + \sigma^2 p} \qquad (9 - 9)$$

وقد تم تعريف الرموز $\sigma^2 p$ و $\sigma^2 i$ و $\sigma^2 e$ في بداية الفصل ويشير الرمز μ إلى المتوسط الإجمالي والذي يحسب من $X_{pJ} = \sum_j \sum_p Xp_i / n_i$ وn_p و Xp_i وهي درجة الفرد p على الفقرة (i). ويمكن أن نبين أن الوسط الإجمالي Xp_i أنه متوسط النسبة المئوية للإجابة الصحيحة للعينة. والمعادلة المناسبة لحساب M (C) هي:

$$M(C) = \frac{^2(c - x_{pl}) + \hat{\sigma}^2 p}{n_i/(\sigma^2 e + \sigma^2{}_i) + {}^2(c - x_{pl}) + \hat{\sigma}^2 p} \qquad (10 - 9)$$

وعلى افتراض أن درجة القطع = 67.، فإن تعويض البيانات في مثالنا يؤدي إلى :

$$0.83 = \frac{^2(0.67 - 0.48) + 0.1015}{0.0273 + {}^2(0.67 - 0.48) + 0.1015} = M(C)$$

ولأن M(c) تتضمن عنصر تباين يعود إلى صعوبة الفقرة ($\sigma^2{}_i$) فإنه يفضل في معامل ليفنجستون. ومن المنطقي طرح استفسار عن العلاقة بين $K^2(X,T)$ و M(C) يمكن أن نبين أن المعادلة (9-8) مكافئة لـ:

$$K^2(X,T) = \frac{^2(c - x_{pL}) + \hat{\sigma}^2 p}{n_i/(\hat{\sigma}^2{}_e + {}^2(c - x_{pl}) + \hat{\sigma}^2 p} \qquad (11 - 9)$$

ومقارنة بالمعادلتين (9-10) و (9-11) نرى انه عند حساب $K^2(X,T)$ و $M(C)$ بناءاً على تحليل التباين الثنائي، فإن الفرق الوحيد بينهما وجود σ^2_i في مقام $M(C)$. وكنتيجة يجب أن تكون $K^2(X,T)$ أكبر من $M(C)$.

مقارنة $K^2(X,T)$ و $M(C)$ و P و *P

أشرنا في البداية أنه يمكن تصور *P أداة لتفسير p. وهذا لا ينطبق على $K^2(X,T)$ أو $M(C)$. فكيف بعدها سيختار مطور الاختبارP من جهة من جهة أخرى $K^2(X,T)$ لتقديم مؤشرات لطبيعة الاختلاف بينهما سنقارن بين P و $K^2(X,T)$ ، كذلك سنقارن بالطريقة نفسها بين P و $M(C)$.

يمكن تحديد المعامل P بشكل اساسي على النحو الآتي: لأي مفحوص صنف باتساق أعط الدرجة أو القيمة 1 وللمفحوصين الذين صنفوا بغير اتساق أعط الدرجة صفر، واجعل Y ترمز إلى المتغير الجديد، وP هي مجموع قيمY مقسومة على أعلى قيمة ممكنة للمجموع والذي يتم الحصول عليه فقطا إذا كانت القرارات جميعها متسقة. وفي حساب $K^2(X,T)$ فإن الدرجة.

$w= (n_ic- x)\ (n_ic- X^{\backprime})$ وهي نفسها للقرارات جميعها، وبسط $K^2(X,T)$ هو متوسط w، والمقام هو أقصى قيمة ممكنة لهذا المتوسط. ولتوضيح الفرق بين $K^2(X,T)$ و P يبين جدول (6 - 9) بيانات افتراضية لثمانية مفحوصين على صيغتين اختباريتين لاختبار مؤلف من 11 فقرة. وكانت درجة القطع لكلا صيغتي الاختبار $V= n_ic$ فقرات. ويبين جدول (6-9): أمثلة على (c- x) $(c\text{-}x^1)$ لاتساق القرار أو عدمه. ويبين الجدول ايضا قيم كل من Y و W. وحيث ان

جدول (6-9): أمثلة على اتساق القرار أو علامة

المفحوص	X	`X	التصنيف	y	$(C\text{-} X^1)\ (\text{-}x)=w$
A	10	11	متسق	1	12
B	8	7	متسق	1	0
C	8	6	غير متسق	0	1-
D	10	2	غير متسق	0	15-
E	5	8	غير متسق	0	2-
F	4	7	غير متسق	0	0
G	5	5	متسق	1	4
H	3	3	متسق	1	16

P تعتمد قيم كل منY على إضافة الدرجة Y فإنه يمكن رؤية أن حسابP : جميع القرارات المتسقة تعطي الوزن نفسه Y =1، والقرارات غير المتسقة جميعها تعطى الوزن نفسهY= صفر، وبما أن $K^2(X,T)$ تعتمد على قيمY فإن الوزن المعطى للقرار يعتمد على اتساق القرار للمفحوص، بالإضافة إلى كم التشابه لدرجات المفحوصين على الصيغتين الاختبارين للمفحوص. على سبيل المثال: تم تصنيف المفحوصين D,C دون اتساق وكانت W للمفحوص D أكثر سلبية منها للمفحوص C، إذ أن الفرق متطرف بين درجتي المفحوص D. وبصورة أساسية فإن هذا يعني أنه عندما يحصل المفحوص على درجات متباعدة مثل 2 و 10، وصنف المفحوص دون اتساق فإن عدم الاتساق هذا يكون جدياً أكثر مما هو لمفحوص درجاته 6 و 8 وكان تصنيفه دون اتساق. ويجب أن يعتمد الاختيار بين P ومعاملات مثل $K^2(X,T)$ أو M(C) على ما إذا وافق الفاحص على القيم المؤشر لها بمعامل القرارات المختلفة.

ولطرح قضية كيفية تأثر $K^2(X,T)$ و M(C) بالتغيرات في درجة القطع، خذ على سبيل المثال: أحد الاختبارات المؤلف من 11 فقرة وتقدم له 8 مفحوصين المستخدم في المثال السابق. يقدم جدول (9–7) مصفوفة افتراضية لاستجابة الفرد– الفقرة على هذه الفقرات، بالإضافة إلى جدول (9–8) الذي يدون كلاً من $K^2(X,T)$ و M(C) لهذه البيانات ولدرجات قطع عديدة، إضافة إلى أن جدول (9–8) يدون تقديرات P و *P لكلا صيغتي الاختبار ولدرجات القطع المختلفة وبشكل عام تشير النتائج إلى أن المعاملات الأربع تتأثر بالطريقة نفسها مع التغير في درجة القطع، وكما بينا في السابق فإن K تتأثر بطريقة معاكسة للتغير بدرجة القطع.

حسابات p و K في التطبيق المفرد:

تطلب حساب P و K في المناقشة السابقة قياسين لكل مفحوص وقد قدم كل من ***(Hunyh,*** ***1976a)*** و ***(Subkovia,1976)***، ***(Peng & Subkoviak, 1980)***، ***(Wilcox ,1981b)*** طريقة لتطبيق واحد في حساب P و K. وطريقة كل من هونيه وسيكوفياك تعتمد على النموذج ثنائي الحد الذي طوره كيتس ***(Keats & Lord, 1962)*** ووصفه لورد ونوفيك (1968) بالتفصيل. ويجب ملاحظة أن كلا الطريقتين لحساب P و K لصيغة افتراضية ومتبادلة مع الاختبار المستخدم تستخدم في عملية جمع البيانات. وكما لوحظ من البداية فإذا كانت الصيغتين متبادلتين فإن درجات الصيغتين يكون لهما التوزيع نفسه والعلاقة نفسها مع درجة النطاق. وتعد التبادلية صيغة أقوى للتكافؤ من التوازي التام، وإحدى مشكلات استخدام معاملات هو نية وسبكوفياك أنه

جدول (9-7) : مصفوفة الفرد- درجة الفقرة لتحليل تباين الفرد - الفقرة لثمانية مفحوصين على 11 فقرة

المفحوص	1	2	3	4	5	6	7	8	9	10	11	المجموع
A	1	1	1	0	1	1	1	1	1	1	1	10
B	1	1	1	0	0	1	1	0	1	1	1	8
C	0	1	1	1	0	1	1	0	1	1	1	8
D	1	0	1	1	1	1	1	1	1	1	1	10
E	0	0	0	1	0	1	1	0	1	0	0	5
F	0	0	0	0	0	0	1	0	1	1	1	4
G	1	0	0	0	0	1	1	0	1	0	0	5
H	0	0	0	0	0	1	1	0	1	0	0	3

مصدر التباين	مجموع المربعات	مصدر التباين	مصدر التباين
المفحوصين	4.715	7	0.673
الفقرات	6.704	10	0.670
البواقي	9.659	70	0.137

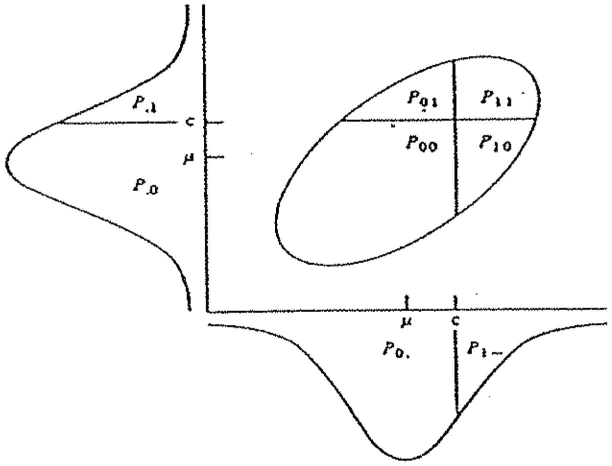
جدول (9-8) : اثر تغير درجة القطع على $\hat{M}(C)$ و $\hat{K}^2(X,T)$، وP و*P

العامل	درجة القطع				
	0.1	0.3	0.5	0.7	0.9
$\hat{M}(c)$	0.94	0.88	0.76	0.76	0.88
$K^2(X,T)$	0.96	0.92	0.83	0.83	0.90
$\hat{P}$	1.00	0.90	70	0.70	0.90
$\hat{P}*$	1.00	0.80	40	0.40	0.80

وفي القياس المحكي يكون اهتمامنا بمعامل اتساق القرار للصيغة قيد الاستخدام وصيغة افتراضية متبادلة مع الصيغة الأصلية، فضلاً عن الاهتمام بشيء مثل قيمة معامل اتساق القرار لجميع الأزواج الممكنة من الاختبارات المبينة بالمعاينة العشوائية من نطاق الفقرات. ما الذي يؤثر في هذه المشكلة؟ بما أن القيمة المتوسطة عبر جميع ازواج الاختبارات أقل من القيمة نفسها لزوج الاختبارات المتبادلة، يبدو أن هذه الطرائق تؤدي إلى تصور إيجابي لكنه زائف (غير منطقي) لاتساق القرار. وقد يكون هذا صحيحاً عندما لا تؤخذ العوامل الأخرى بعين الاعتبار ومشكلة أخرى ممكنة مع هذه الطرائق تظهر عندما يستجيب المفحوصين جميعهم على الفقرات نفسها، فإن استخدام نموذج ثنائي الحد يقتضي التساوي في صعوبة الفقرات – وهذا متطلب لا يتحقق عملياً– وفشل تحقيق هذا الفرض يعني أن الطرائق تؤدي إلى قيم أقل من المحسوبة لاتساق القرار لأي زوج من الاختيارات المتبادلة (Hunyh& Saunders, 1980) لذا فإن أثر هاتين المشكلتين تلغي إحداهما الأخرى.

وفي طرائق سبكوفياك وهونية، فإن الأخيرة طورت بدرجة أكبر، فهي تسمح بتقدير الأخطاء المعيارية لكل من P و k. وكلا الطريقتين معقدة رياضياً لذا فلن نتعرض للتفاصيل الرياضية، وتتوافر قائمة من البرامج المحوسبة لاستخدام طريقة هونية ومبينة في (Hunyh, 1980a). وتتضمن هذه الدراسة أيضاً جداول يمكن استخدامها في تحديد قيم p و K لاختبارات عدد فقراتها ما بين 5–10. ولاستخدام هذه الجداول فمن الضروري معرفة μ، kR21 ، ودرجة القطع فقط. ويستخدم KR21 على أنه ثبات درجات الاختبار الذي يتطابق مع النموذج ثنائي الحد. والنتائج التي دونها سبكوڤياك (1978) تقترح أنه عند الاعتماد على عدد قليل من المفحوصين مثل العدد 3، فإن طريقة هونية توفر تقديرات دقيقة عادلة لـ P وK للاختبارات المتوازية تماماً وتحتوي 10 فقرات على الأقل.

وقدم كل من بينج وسبكوڤياك (1980) تقريب بسيط لعمل هونيه وغير متحيز (عادل). خذ على سبيل المثال عند توافر مجموعتين من القياسات التي لها توزيعات اعتدالية متماثلة ولها توزيع طبيعي ثنائي متصل Joint bivariate normal distribution. فمثل هذه الاختبارات تعد متبادلة، وهذه الحالة موضحة في الشكل (9–2). وعندما تكون توزيعات درجات الاختبار متشابهة يمكننا أن نبين أن :



شكل (9-2) التوزيع الاعتدالي ثنائي الحد.

$P = 2(P_{o.} - P_{oo}) + 1$، لذا فان

$P^* = 4(P_{o.} - P_{oo}) + 1$، كذلك يمكن أن نبين ان:

$K = P_{o.} - P_{oo} \div P_{o.} - P^2_{o.}$ ، والان $P_{o.} = Pr\ (n_i c > x)$: احتمالية كون X اقل من $n_i c$ على افتراض ان الدرجات موزعة عشوائياً، لذا يمكن حساب .po باستخدام :

$$P_{o.} = Pr \left\{ \frac{\mu_x - n_i c}{\sigma x} > Z \right\} \qquad \text{.....................(12- 9)}$$

حيث تمثل Z متغير طبيعي معياري عشوائي، فالحد على الطرف الأيسر هو المسافة على يسار $\frac{n_i(C- \mu_x)}{\sigma_x}$ في التوزيع الطبيعي. للتوضيح، إذا كانت $\mu_x = .o$، و $\sigma_x = 10$، و $n_i c = 60$ فان $po. = Pr\ \{ n_i c > x\} = Pr\ \{ \frac{\mu_x - n_i c}{\sigma_x} > Z\}$

$$Pr = \{\frac{50-60}{10} > Z\} = 0.84$$

والآن، فإن Poo هي احتمالية كون الدرجات على كلا الصيغتين اقل من $n_i c$، أو بشكل بديل أن كلاً المتغيرين الطبيعيين المعياريين العشوائيين اقل من $\frac{\sigma_x}{n_i c - \mu_x}$ ، وهذه الاحتمالية يمكن إيجادها من جداول التوزيع الطبيعي المعياري ثنائي المتغير (Gupta,1963) شريطة كون الارتباط بين المتغيرين معروف. وهنا يكون الارتباط هو معامل الثبات المعرف بنظرية القياس التقليدية، لذا يمكن حسابه بالتطبيق لمرة واحدة واستخدام أية طريقة مناسبة في

حسابه. ولتقريب طريقـــة هونيه يستخدم KR21. ويبين جدول (9-9) جدول مختصـــر لـ p كدالة لـ Z، حيث $Z = (\mu x - n_i c)/\sigma_x$، و P هي الارتباط بين الصيغتين. ولتوضيح حساب كلا من *p و k أفترض أن اختبار مؤلف من 7 فقرات، $\mu x = 4.5$، و $\sigma_x = 1$، $KR21 = 0.4$، $n_i c = 5$ بعد ذلك وباستخدام جدول الدرجات الرائية، فإن:

$$Po = Pr\{\frac{\hat{\mu}_x - n_i c}{\hat{\sigma} x} > Z\} = Pr\{\frac{4.5 - 5}{1} > Z\}$$

$$pr(0.5 > z) = 0.69$$

جدول (9-9): احتمالية ارتباط متغيرين اعتدالين معياريين الارتباط بينهما أقل أو يساوي Z.

Z	p=30	p=30	.40	.50	.60	.70	0.80
-1.00	.0455	.0455	.0536	.0625	.0725	.0840	.0976
-0.90	.0578	.0578	.0671	.0773	.0887	.1015	.1167
-0.80	.0726	.0726	.832	.047	.1073	.1216	.1383
-0.70	.0902	.0902	.1020	.1147	.1286	.1442	.1625
-.060	.1106	.1106	.1237	.1376	.1527	.1696	.1893
-0.50	.1342	.1342	.1483	.1633	.1796	.1976	.2186
-0.40	.1609	.1609	.1760	.1920	.2092	.2282	.2503
-0.30	.1908	.1908	.2067	.2235	.2415	.2614	.2843
-0.20	.2239	.2239	.2404	.2577	.2763	.2968	.3204
-0.10	.2595	.2595	.2767	.2944	.3134	.3343	.3583
0.00	.2985	.2985	.3155	.3333	.3524	.3734	.3976
0.10	.3395	.3395	.3564	.3741	.3930	.4139	.4379
0.20	.3824	.3824	.3989	.4162	.4348	.4553	.4789
0.30	.4266	.4266	.4426	.4593	.4773	.4972	.5202
0.40	.4718	.4718	.4869	.5028	.5200	.5391	.5612
0.50	.5171	.5171	.5312	.5462	.5625	.5805	.6015
0.60	.5621	.5621	.5752	.5891	.6042	.6211	.6408
0.70	.6062	.6062	.6181	.6308	.6447	.6603	.6786
0.80	.6489	.6489	.6595	.6710	.6836	.6979	.7146
0.90	.6896	.6896	.6990	.7092	.7205	.7334	.7486
1.00	.7282	.7282	.7363	.7452	.7552	.7667	.7803
1.10	.7640	.7640	.7709	.7787	.7874	.7975	.8096
1.20	.7970	.7970	.8028	.8094	.8169	.8257	.8363
1.30	.8269	.8269	.8318	.8343	.8438	.8513	.8605
1.40	.8538	.8538	.8578	.8624	.8678	.8742	.8821
1.50	.8777	.8777	.8809	.8847	.8892	.8946	.9012
1.60	.8987	.8987	.9012	.9043	.9079	.9124	.9180
1.70	.9168	.9168	.9188	.9212	.9242	.9279	.9325
1.80	.9324	.9324	.9339	.9358	.9382	.9411	.9449
1.90	.9455	.9455	.9467	.9482	.9500	.9521	.9555

وللحصول على p_{00} في مثالنا نستخدم جدول (9–9) وباستخدام $Z = c\hat{n}_i - \mu \,\hat{/}\, \sigma_x$

$$= \frac{4.5 - 5}{1} = 0.5 \text{ ، و } KR21 = 0.4$$

وباستخدام جدول (9–9): $P_{00} = 0.053$ ، فان

$$P^* = 1 + 4\,(0.53 - 0.69) = 0.36$$

$$\text{و } k = \frac{0.53 - 0.69^{\,2}}{0.69 - 0.69^{\,2}} = 0.25$$

دقة القرار:

أحد الاستخدامات الرئيسة للاختبارات محكية المرجع هو اتخاذ قرارات تتعلق بمستويات كفاءة المفحوصين ففي العديد من الولايات يعطى طلبة المدارس العليا الدبلوما بعد اجتياز درجة محددة في اختبار الحد الأدنى من الكفاءة، Minimum Competency Examination، وفي المعاهد التربوية العليا يتم تسكين الطلبة في صفوف عادية أو علاجية أو متميزة اعتماداً على درجاتهم في اختبار محكي المرجع. فعندما نستخدم درجات الاختبار بهذه الطريقة فإن أحد مقاييس نوعية الاختبار هو دقة القرارات المأخوذة على أساس هذه الدرجات (كتمييز لها عن اتساق القرارات الناجمة عن تكرار العملية الاختبارية). أفترض أننا نعرف قيمة μ_p درجة النطاق لكل مفحوص، نحدد بعدها درجة القطع على متصل درجة النطاق، ويعد هنا كل من كانت درجته تساوي C أو أكبر منها مسيطراً على مجال المعرفة، والذين درجاتهم أقل منها يعدوا غير مسيطرين على المجال أفترض أيضاً أن الاختيار محكي المرجع طبق على المفحوصين جميعهم لتصنيفهم إلى فئتين الأولى مسيطرين والأخرى غير مسيطرين، فالمفحوصين الذين درجاتهم عند درجة القطع C أو أعلى منها هم المسيطرين والذين درجاتهم أدنى منها هم غير المسيطرين، ويمثل الشكل (9–3) مثل هذا القرار. (الفرق بين الشكلين (9–1) و (9–3) هو أن الأسبق يشير إلى العلاقة بين القرارات المأخوذة باستخدام درجات اختبارية ملاحظة عددها اثنتان ومختلفتان، والأخير يشير إلى العلاقة بين القرارات المأخوذة باستخدام درجة اختبار واحد، وهذا يتم إجراؤه فيما لو تمت معرفة درجات النطاق).

وبالتمحيص لشكل (9–3) يتبين لنا أن هنالك نوعية من القرارات الصائبة ونوعين من

القرارات الخاطئة (غير الصائبة) المحتملة لكل مفحوص. فالقرارات الصائبة هي القرارات الصحيحة الموجبة والصحيحة السالبة. فالقرار الصائب الموجب يظهر عندما يصنف المفحوص المسيطر من خلال درجة النطاق ويكون التصنيف صحيحاً، يرمز لاحتمالية القرار الصائب الإيجابي بـ P_{11}. . بينما القرار الصائب السلبي يظهر عندما يكون المفحوص غير مسيطر من خلال درجة النطاق ومصنف غير مسيطر على اساس درجة الاختبار، واحتمالية هذا القرار يرمز لها بـ $\hat{p}_{00}$. وبتفحص شكل (9–3) يظهر أيضاً أن هنالك نوعين من

التصنيف بناء على درجة النطاق

	غير مسيطر	مسيطر	
$C >$ غير مسيطر تصنيف بناء على درجة الاختبار	قرار صائب سلبي $0.4 = p_{00}$	قرار خاطئ سلبي $0.2 = p_{01}$	$0.6 = p_{0\cdot}$
$C <$ مسيطر	قرار خاطئ ايجابي $0.1 = p_{10}$	قرار صائب ايجابي $0.3 = p_{11}$	$0.4 = p_{10}$
	$0.5 = p_{\cdot 0}$	$0.5 = p_{\cdot 1}$	

شكل (9-3): احتمالية التوافق بين تصنيف درجة النطاق ودرجة الاختبار.

القرارات غير الصائبة (الخاطئة) قد تظهر. فالقرار الخاطئ الإيجابي يظهر عندما يصنف المفحوص غير المسيطر من خلال درجة النطاق على انه مسيطر من خلال درجة الاختبار، ويرمز لاحتمالية هذا القرار بالرمز $\hat{p}_{10}$ والقرار الخاطئ السلبي يظهر عندما يكون المفحوص مسيطر من خلال درجة النطاق ويصنف على انه غير مسيطر من خلال درجة الاختبار، وهذه الاحتمالية يرمز لها $\hat{P}_{01}$ ومن الواضح أن المعلومات المعطاة من خلال هذه الاحتمالات تفيد في تقويم الاختبار محكي المرجع. على سبيل المثال $\hat{P} = \hat{P}_{00} + \hat{P}_{11}$ ($\hat{P}$: احتمالية القرار الصائب)، ويعد مقياس طبيعي لعدم دقة القرار. ولبيانات شكل (9-3)، فإن:

$P = 0.4 + 0.3 = 0.7$ ، لذا فإن فإن 70% من القرارات تكون صائبة.

ولتقويم درجة دقة القرار لاختبار معين بدرجة قطع معينة، يجب حساب احتمالية ظهور كل

من القرار الخاطئ الإيجابي، والقرار الخاطئ السلبي. وقد قدم ويلكوكس(Wilcox,1977a) عدة طرائق لحساب P_{10} و P_{01} : احتمالات أخطاء التصنيف: الخطأ السلبي والخطأ الإيجابي على التوالي، وتستخدم إحدى هذه الطرائق النموذج ثنائي الحد. وقد زودنا هونيه (Hunyh, 1980d) بجداول يمكن استخدامها لتطبيق هذه الطريقة لاختبارات مؤلفة من فقرات عددها من (5–10)، وكذلك زودنا ببرنامج محوسب يستخدم هذه الطريقة لحساب تقديرات P_{10} و P_{01} لاختبار يصل عدد فقراته إلى60 فقرة، وطور ويلكوكس طريقة أخرى يمكن إجراؤها يدوياً شريطة استخدام جداول تحويلات الاقترانات الدائرية (ARC SIN) وجداول التوزيع الطبيعي الثنائي (وهذه الطريقة تجدها واضحة في الجزء الرابع من (wilcox, 1977a) ومع أ كلا الطريقتين طورتا لحساب P_{10} و P_{01} إلا أنه يمكن استخدامها في حساب P_{10} و P_{00}. وستجد مناقشة إضافية إثرائية لطرائق تقويم دقة القرار للاختبارات محكية المرجع في الفصل 18(اعداد معايير الاختبار).

الخلاصة:

نطاق الفقرات هو مجتمع الفقرات التي يجب أن يتقدم إليها المفحوص عندما يطبق عليهم اختبار معد ليطابق مواصفات معينة. ويمكن التفكير بدرجة النطاق للمفحوص على أنها نسبة الفقرات من النطاق التي يستطيع المفحوص الاجابة عليها إجابة صحيحة. وتستخدم نتائج الاختبارات محكية المرجع إما لتقدير درجة النطاق أوموقف السيطرة. وكما هو مبين في هذاالفصل يستخدم تقدير درجة النطاق في تفسير درجة الإجابات الصحيحة في الاختبار على أنها تقدير لدرجة النطاق. ولتحديد حالة السيطرة تقارن درجة الإجابات الصحيحة بدرجة قطع مناسبة محددة مسبقاً فإن كانت درجة الإجابات الصحيحة أكبر من درجة القطع أو تساويها فإن المفحوص يصنف مسيطراً على النطاق، عدا ذلك فإنه يصنف غير مسيطر على النطاق.

ويرتبط بهذين الهدفين للاختبارات المحكية مشكلتي ثبات. الأولى تتعلق بالمدى الذي تقترب فيه درجة الاجابات الصحيحة الملاحظة من درجة النطاق. والأخرى تتألف من مشكلتين فرعيتين، إحداها تتمثل في موقف السيطرة المبنى على

أساس القياسين، والثانية تتمثل في دقة موقف السيطرة، أي مدى اتفاق موقف السيطرة المبني على درجة الاختبار الملاحظة مع الموقف الممكن فيما لو كانت درجة النطاق معروفة.

وعندما يتم بناء اختبار بأسلوب المعاينة العشوائية للفقرات فإن معامل إمكانية التعميم(ρ_L^2) وتباين الخطأ $\left(\frac{\sigma_i^2+\sigma_e^2}{n_i}\right)$ يمكن استخدامهما في تكميم دقة تقدير درجات النطاق. ومن هاتين القيمتين يفضل استخدام تباين الخطأ على معامل إمكانية التعميم، وذلك لأنه يعتمد فقط على دقة التقدير، في حين أن معامل إمكانية التعميم يعتمد على تباين درجات النطاق وعلى دقة تقدير درجة النطاق. وبالطبع فإن المقياس الأبسط لاتساق قرار السيطرة هو (p) أي نسبة المفحوصين المصنيفين باتساق على أنهم مسيطرين أو غير مسيطرين باستخدام مقياسين اثنين وكلاهما محكي المرجع.

ويوجد تحويلين لـP يمكن استخدام أي منهما وهما *P و k لكوهين. ويمكن تصور كل منهما على أنه تحويل لـP إلى تدريج جديد تكون فيه نقطتي التدريج الصفر والواحد مفسرتين (قابلتين للتفسير). وتتأثر قيمة P وتحويلاتا بعوامل عدة، مثل امكانية التعميم من الدرجات الملاحظة إلى درجات النطاق، وطول الاختبار، وموقع درجة القطع، والتشابه في توزيعي الدرجات الملاحظة لموقفي القياس. وباستخدام النموذج ثنائي الحد من الممكن حساب P و K من تطبيق واحد للاختبار. واقترح كل من بينج وسيكوفياك تقريباً لهذا التقدير.

ويوازن المعامل (P) القرارات المتسقة والقرارات غير المتسقة جميعها بالتساوي، لذا فإنه لا يعكس الاتساق وعدمه في تعيين الإفراد في تصنيفات السيطرة، لذلك توجد معاملات أخرى تعكس درجة الاتساق وعدمه. اثنان من هذه المعاملات هما $K^2(X,T)$ لليفنجستون و M(C) لبرينان وكين، ويمكن حساب $K^2(X,T)$ من تطبيق الاختبار مرة واحدة أو مرتين على حد السواء. والأكثر شيوعاً هو حساب M(C) من تطبيق الاختبار لمرة واحدة وقيمة $K^2(X,T)$ أكبر من قيمة M(C)، وذلك لأن عنصر الخطأ المحدد في M(C) يتضمن تنوعاً يعود إلى صعوبة الفقرة.

بالإضافة إلى الاهتمام باتساق القرار الذي قد ينتج عن القياس محكي المرجع، هناك أيضاً نوع آخر من الاهتمام يستخدم دقة القرار. وفي دقة القرار تحدد أربعة أنواع من القرارات:

الصح الإيجابي- يصنف المسيطر على أنه مسيطر، والصح السلبي - يصنف غير المسيطر على أنه غير مسيطر، والخطأ الإيجابي - يصنف غير المسيطر على أنه مسيطر، والخطأ السلبي- يصنف المسيطر على أنه غير مسيطر. ولتقويم درجة دقة القرار لاختبار معين له درجة قطع معينة فيجب حساب احتمالات ظهور كل من الخطأ الإيجابي والخطأ السلبي. وقد وصف كل من ويلكوكس وهونيه طرائق يمكن استخدامها في هذه الحسابات.

التمارين:

1/ اعتمد مصفوفة الفقرة – المفحوص أدناه في الإجابة عن الأسئلة التي يليها:

المفحوص	الفقرات							
	1	2	3	4	5	6	7	8
1	0	0	1	1	0	1	1	1
2	1	1	0	1	1	1	1	0
3	1	1	1	1	1	1	1	0
4	0	0	0	1	0	0	1	0
5	1	0	1	0	1	0	0	1
6	1	0	1	1	1	0	1	0

أ) ما خطأ التباين المحسوب لمجموعة ا لبيانات وباستخدام الصيغة (9–1).

ب) ما الخطأ المعياري للقياس.

ج) ما اقصى قيمة محتملة للخطا المعياري للقياس لهذه البيانات.

2/ أجرى باحث تحليل تباين هويت لاستجابات الفقرة المبينة في التمرين الأول للحصول على قيمة معامل ألفا (KR20). والنتائج مدونة أدناه.

مصدر التباين	مجموع المربعاتSS	درجات الحرارة	متوسط المربعات
المفحوصين	1.854	7	0.370
الفقرات	1.645	10	0.235
البواقي	7.979	35	0.222

أ) ما معامل إمكانية التعميم لحساب قيم درجات النطاق لهؤلاء المفحوصين.

ب) هل يؤشر هذا المعامل إلى أن درجات النطاق تمثل بدقة درجة الإجابة الصحيحة الملاحظة على هذا الاختبار.

ج) إذا كان هذا المعامل =1، فما درجة النطاق الصحيحة للمفحوصين 1، 2، 3.

د) افترض أن الفاحص اهتم بالمفاضلة عبر المفحوصين من خلال درجات نطاقهم. ما صيغة معامل إمكانية التعميم المناسب بدلالة متوسط المربعات.

هـ) ما قيمة معامل إمكانية التعميم المحسوب باستخدام الصيغة المستخدمة في الفرع د؟

و) هل يمكن حساب هذه القيمة بطريقة أخرى؟ ما هي؟

3/ افترض البيانات أدناه التي تشير إلى الدرجة الكلية التي حصل عليها المفحوصين في صيغتين لاختبار مستوى الكلية، ومؤلف من 40 فقرة لكل صيغة، وعلى افتراض أن المشرف حدد أربع درجات نجاح ممكنة لهذا الاختبار هي: 60% صح، 70%صح، 80%صح، 90% صح. فأجب عن الأسئلة الآتية:

أ) حدد نسبة المفحوصين الذين صنفوا باتساق من كلا الصيغتين باستخدام كل درجة نجاح مقترحة؟

ب) ما قيمة معامل كابا (k) لكلا صيغتي الاختبار ولكل درجة قطع مقترحة؟

ج) ما قيمة p^* لكلا صيغتي الاختبار باستخدام كل درجة قطع مقترحة؟

د) لقيم كل من P، k، P^* وعند درجة قطع60%، الكتب تفسيراً لمعنى الإحصاءات الثلاثة هذه؟

الاختبار الثاني	الاختبار الأول	المفحوص
22	21	1
23	25	2
23	36	3
29	26	4
35	23	5
30	26	6
38	35	7
38	38	8
24	34	9
34	24	10
29	28	11
28	29	12
27	27	13
28	28	14
32	25	15
29.33	28.33 =	$\hat{\mu}$
5.05	4.94 =	$\hat{\sigma}$
	0.20 =	$\hat{p}$

هـ) ما قيمة معامل ليفنجستون المحسوب لكلا صيغتي الاختبار عند درجة قطع60%؟

4/ أ- عد إلى مصفوفة استجابة المفحوص- الفقرة المبينة في التمرين الأول والثاني. ما قيمة معامل ليفنجستون لهذا الاختبار عند درجة قطع 0.50؟ وعند درجة قطع 0.75؟

ب- بدون إجراء حسابات تنبأ كيف تكون قيم دليل الثبات لبرنيان – كين لهذه البيانات مقارنة بتقديرات ليفنجستون؟ فسر تنبؤك؟

جـ- احسب دليل برنيان – كين لهذه البيانات باستخدام درجات قطع 0.50 ، 0.75 ، على التوالي.

د- راجع صيغة برنيان – كين معادلة 9–10، وبين أن كانت تزودنا بدليل لدقة القرار أو بحساب دقة درجة النطاق عدل استجابتك.

هـ- في اي الحالات يمكن اعتبار صيغة برنيان- كين على انها دليل «معياري المرجع».

5/ أ- لأهداف هذا التمرين أفترض أن الإحصاء الوصفي للاختبار (1 في تمرين 3) غير متوافر، إضافة إلى افتراض أن $\hat{P} = 0.4$ احسب قيمة k لدرجة قطع النسبة الصحيحة 0.75.

ب- ما الافتراضات الضرورية اللازمة لطريقة الحساب التي اتبعتها.

6/ في وصف اختبار الكفاءة القرائية الاساسي دون جانا بول (*Gana pole, 1980*) لاختبار محكي مؤلف من اختبارات فرعية عددها ستة يتألف كل منها من 10 فقرات.

لتأكيد الثبات لدرجة أكبر لاختبار الكفاءة القرائية الأساسي استخدمت طريقة اتساق القرار لمقارنات كل من: الاختبار – الإعادة، والصيغة المكافئة. وحسب دليل لكل من نسبة الاتفاق بالنسبة لمستويات الكفاءة الثلاثة: 80%و 70% و 60%، وقيم اتساق قرار الاختبار الإعادة للاختبارات الفرعية وتراوحت بين 27% إلى 88% للصيغة (أ) ومن 37% إلى90% للصيغة (ب).

أ- في المقالة نفسها، دون المؤلف تقديرات ثبات الاختبار – الإعادة على الصيغة (أ) وتراوحت بين 50 ألى 0.71 لماذا كانت هذه التقديرات منخفضة بهذا الشكل.

ب- ما الأثر المحتمل لاتساق قرار الاختبار – الإعادة فيما لو ازداد طول الاختبارات الفرعية جميعها بخمس فقرات.

ج- افترض أن الصيغة (أ) للاختبار طبقت على مجموعة مفحوصين درجات نطاقهم متجانسة جداً. كيف يتأثر دليل اتساق القرار لهذا العينة.

د- افترض أن مطور الاختبار بدلاً من ذلك دون اتساق القرار بدلالة *p و k، فكيف يمكن مقارنة هذه القيم بتلك القيم المدونة؟

هـ- دون مطور الاختبار أن الدرجات عليهذا الاختبار مقيدة في المدى ويتجمع معظم الدرجات عند النهاية العليا لمدى الدرجات المحتمل لكلا الموقفين في هذه الظروف لأي درجة قطع يكون دليل اتساق القرار أعلى؟

الوحدة الثالثة

الصدق

الفصل العاشر

10

مقدمة الى الصدق

الفصل العاشر

مقدمة الى الصدق

تستخدم الدرجات عادة للحصول على استنتاجات عن سلوك المفحوصين فيما وراء العملية الاختبارية. فمثلاً في اختبار علوم للصف الثامن، ومقياس التمركز حول الذات لطفل مرحلة ما قبل المدرسة، واختبار المحاماة للولاية، يهدف الفاحص في الحالات جميعها الحصول على استدلال من خلال درجة الاختبار على نطاق أكبر من السلوكات التي يمكن أن يظهرها المفحوص الآن أو في المستقبل. ويتطلب الاستخدام المسؤول لدرجات الاختبار من مستخدمها أن يكون قادراً على تعديل الاستدلالات التي حصل عليها، وذلك بأن يكون لديه حجة مقنعة لاستخدام درجات الاختبار للهدف الحالي، ولاختياره لهذا الاختبار وتفضيله على طرائق التقييم الأخرى المتوافرة (Mĕssick, 1981). ومثل هذه التعديلات لها متطلبين سابقين هما: ثبات درجات الاختبار وصدقها. فعامل الثبات الكبير يشير إلى اتساق درجات المفحوصين، ولكنها لا تؤكد استدلالات عن المفحوصين يمكن الدفاع عنها. أفترض مثلاً التشابه الجزئي في خزان وقود السيارة والذي يسجل بانتظام 4/1 زيادة عن المستوى الفعلي للوقود في الخزان. وعندما تؤخذ قياسات متكررة تحت نفس الشروط فإن القياسات تكون متسقة (أي ثابتة)، ولكن هذا الاستدلال عن الوقود والخزان خاطئ، لذا نرى أن دليل ثبات القياس قد لا يكون كافياً لتعديل الاستنتاج المرغوب.

وقد وصف كرونباخ (Cronbach, 1971) الصدق على أنه العملية التي يجمع مطور الاختبار أو مستخدمه من خلالها الأدلة التي تدعم الاستدلالات التي استخلصها من درجات الاختبار. وللتخطيط لدراسة الصدق يجب تحديد الاستنتاج المرغوب بوضوح ومن ثم تصميم دراسة تجريبية لجمع الأدلة اللازمة لملائمة وفائدة الدرجات لمثل هذا الاستنتاج. وهنالك ثلاثة أنواع من الدراسات الرئيسية للصدق هي:

1) صدق المحتوى **Content Validation**: فعندما يرغب مطور الاختبار الوصول إلى استنتاجات من خلال درجة المفحوص على نطاق أكبر من الفقرات المشابهة لفقرات الاختبار نفسه.

2) الصدق المرتبط بمحك **Criterion related Validation:** ويستخدم في المواقف التي يرغب مستخدم الاختبار من خلالها الحصول على استنتاجات واستدلالات من درجة المفحوص في الاختبار عن سلوكاته لمتغير حقيقي له أهمية عملية.

3) صدق البناء : **Construct Validation** وهذا عندما لا يكون هنالك نطاق أو محك متفق عليه مناسب للمحتوى لتحديد نوعية ما يراد قياسه (Cronbach - S Mehi,1955) ويرغب مطور الاختبار إجراء استدلالات من خلال درجات الاختبار على الأداء الذي يمكن تجميعه تحت اسم بناء نفسي معين.

وعند اختيار اختبار لهدف معين، فعلى مستخدم الاختبار التأكد من أن له أدلة صدق مناسبة للاستخدام المقصود في هذا الموقف، وعندما لا يقدم ناشر الاختبار أدلة صدق للاختبار ويعتقد مستخدم الاختبار أن الاختبار له فائدة ممكنة فعليه إجراء دراسة صدق محلية قد تصمم وتُجرى من قبل مستخدم الاختبار. كذلك فإن أنواع دراسات الصدق المختلفة تدعم أنواع الاستدلالات المختلفة لذا، يجب اعتبارها متبادلة. وأخيراً فإن تعديل بعض الاستنتاجات يتطلب أكثر من نوع واحد من دراسات الصدق.

صدق المحتوى:

تهدف دراسة صدق المحتوى تقييم ما إذا كانت الفقرات تمثل نطاق الأداء أو البناء المستهدف بشكل مناسب. فمثلاً، في اختبار الألفاظ نادراً ما يهدف مستخدم الاختبار ما إذا كان المفحوص يعرف معاني كلمات محددة فقط، ولكنه يهتم بمعرفة المفحوص "بالكلمات المشابهة لها". وفي صدق المحتوى هنالك طريقة معتادة هي تحكيم مجموعة من الخبراء المستقلين (غير كتبة الفقرات) للحكم على ما إذا كانت عينة الفقرات مناسبة لنطاق القياس ويدّعي مطور الاختبار أحياناً أن كتابة الفقرات من نطاق محدد بدقة يؤكد صدق المحتوى، ولكن هذه لا تكوِّن بالفعل دراسة صدق محتوى. فصدق المحتوى عبارة عن سلسلة نشاطات تأخذ مجراها بعد التشكيل الأولي للأداة المراد تطويرها، وهذه الأنشطة قد يجريها مطور الاختبار. ويتضمن صدق المحتوى على الأقل الخطوات الآتية:

1) تحديد نطاق الأداء المستخدم.
2) اختيار فريق من الخبراء المؤهلين في مجال النطاق.
3) التزويد بهيكل بنائي لعملية مطابقة الفقرات لنطاق الأداء.
4) جمع تلخيص البيانات الناتجة عن عملية المطابقة.

ولأن صدق المحتوى يستخدم عموماً مع اختبارات التحصيل[1] فإن نطاق الأداء يحدد عادة

---

(1) . في السبعينات كان هنالك اهتماماً ملحوظاً في صدق المحتوى للاختبارات المستخدمة في اختيار المستخدمين، وهنالك دراسات مهمة طورت في هذا السياق، فقد ترغب في مراجعة بعضها، مثلاً (Brien, 1977) و (Lawsh, 1975) و(Gavin, 1977) وذلك لمناقشة صدق المحتوى في سياق اختبارات الشخصية.

من خلال قائمة الأهداف التعليمية. ومع ذلك فنحن نميز الطرائق الأخرى المستخدمة في تحديد النطاق (انظر فصل 4). وسيستخدم مصطلح الأهداف في الأجزاء التالية على مدى واسع ليرمز إلى أي قائمة من السلوكات أو فئات السلوك التي يستخدمها مطور الاختبار أو مستخدمه في وصف النطاق.

اعتبارات عملية في صدق المحتوى:

يتطلب التخطيط لدراسة صدق المحتوى القرارات العملية الآتية:

1) **هل تم موازنة الأهداف لتعكس أهميتها؟** إحدى الطرق الشائعة هي افتراض قيمة أو وزن متساو للأهداف المدرجة جميعها. وجهة النظر هذه غير شاملة، واقترح كاتز (Katz, 1958) وزن أو ترتيب الأهداف حسب أهميتها قبل المزاوجة بين الأهداف والفقرات. واقترح كلين وكوزكوف (Klein & Kosecoff, 1975) تقدير أهمية كل هدف على متصل مؤلف من خمس نقاط ودون الحاجة إلى تقديرات مطلقة من قبل الخبراء الذين حكموا الفقرات في عملية الموازنة، كمثال في مدارس المقاطعات الكبيرة يمكن الحصول على الأوزان النسبية للأهداف المختلفة من خلال اشتراك معلمي المقاطعة جميعاً، مع أن عينة أقل بكثير اشتركت فعلاً في مهمة المطابقة بين الفقرة والهدف، ونقطة أخرى يجب مراعاتها هي طبيعة التعليمات التي أعطت للخبراء الذين طلب منهم تقدير أهمية الأهداف. كمثال في الحكم على أهمية الأهداف التدريسية فقد يأخذ أحد المدرسين الوقت المستغرق في التدريس محك، وآخر قد يحدد الأهمية النسبية من خلال العلاقة بالتحكم المستقبلي. وفي صدق المحتوى يجب على مصمم الدراسة أن يزود الخبراء بتعريف مشترك متفق عليه للأهمية لا استخدام الأحكام الخاصة الذاتية للمحكمين (الخبراء).

2) **كيف يمكن تكوين مهمة المزاوجة للفقرات مع الأهداف؟** أسلوب فعال يكمن في إرشاد الخبراء للتقدم بانتظام عبر الاختبار بمزاوجة كل فقرة إلى قائمة الأهداف. وقد اقترح كل من كاتز (Katz, 1985) وايبل (Ebel, 1956) أن يقرأ الخبير الفقرات قراءة تصورية وتحديد الاستجابة كما يفعل المفحوص. وقد قدم كلين وكوزكوف (1975) ثلاثة مقترحات إضافية لتسهيل مهمة المزاوجة وهي:

- كتابة الفقرات كل واحدة على بطاقة منفصلة.
- مقارنة كل فقرة إلى قائمة الأهداف التعليمية.
- تسجيل نتاج قرار المزاوجة بصيغة معيارية.

ترتيب المزاوجات حسب درجة القوة

كذلك وصف هامبلتون (Hambelton, 1980) طريقة تقدير درجة المزاوجة بين الفقرة والهدف المحدد باستخدام تدريج خماسي النقاط تشير فيه النقطة (1) إلى درجة مزاوجة ضعيفة (5) إلى درجة مزاوجة ممتازة. ويحسب وسط أو وسيط التقديرات المزاوجة الكلية بين الفقرة والهدف.

3) **ما هي مظاهر الفقرة التي يجب فحصها؟** يجب تزويد الحكام بوصف واضح لخصائص الفقرة والنطاق ليأخذها بعين الاعتبار في عملية مزاوجة الفقرات مع النطاق. وبعض الخصائص المهمة: الموضوع، والعمليات العقلية أو مستوى تعقيد الأداء المطلوب، بالإضافة إلى نموذج الاستجابة المطلوب. وعلى سبيل المثال خذ أهداف الرياضيات الآتية للصف الأول:

أ – جمع أي عددين موجبين مجموعها 18 أو أقل.

ب– طرح أي عددين (قيمة كل منهما أقل من 20) وناتج الطرح قيمة موجبة. ويحتوي الشكل (10–1) عينة فقرات يمكن أن تستخدم لتقييم الأداء في هذا النطاق. فالفقرة (1) تتزاوج مع الهدف أ، والفقرات (2، 3) تتزاوج مع الهدف ب، والفقرة (4) تتعلق بموضوع ليس من محتوى الهدفين، والفقرة تتطلب مستوى أداء أعلى من المتضمن بالأهداف إذ أنها تتطلب الجمع بين المهارات المطلوبة للهدفين المنفصلين لحل مشكلة واحدة.

وحتى عندما يكون محتوى الفقرة والعملية مناسبين للنطاق، فإن أسلوب أو شكل العرض قد يؤدي إلى استفسار الخبراء عما إذا كانت الفقرات تطابق المجال أو النطاق. فعلى سبيل المثال يمكن اعتبار الفقرات (من 1 إلى 3) في الشكل (10– 1) غير ملائمة لمنهاج تدريس فيه مشكلات الجمع والطرح بصيغة عمودية وليست خطية كما المبينة في الشكل. وأحياناً يتطلب نموذج العرض من المفحوصين تمثيل سلوكات إضافية خارج النطاق المستهدف، مثل سؤال 6 في الشكل (10– 1)، إذ يتطلب قدرة المفحوصين على القراءة فيما لو تم حل المسألة بشكل صحيح. ومثل هذه الفقرة لا تمثل نطاق مهارات الجمع والطرح الأساسية بوضوح. وأخيراً يجب أن يؤخذ بعين الاعتبار طبيعة الاستجابة المطلوبة من المفحوص، فاختبار القراءة والكتابة يبدو أنها تمثل نطاقات أداء مختلفة نوعاً ما، وذلك لأن طبيعة الاستجابة المطلوبة مختلفة مع أن الفقرات (أو المثيرات) تبدو متشابهة في كلا الموقفين.

حالما يتم بناء الاختبار فإنه يجب سؤال الخبراء عن خصائص الفقرات (نوقشت في الفصل 4)، والحكم على مدى تحقيق الفقرات لهذه الخصائص. ومع أن هذه المهمة غير متحيزة إلا أنها تتضمن المرحلة الأولى فقط من عملية ثنائية المرحلة، ومن الضروري أيضاً

1- 3 + 5= ....

2- 12 - 10= ....

3- 8 - 5 = ...

4- 25 - 16 = ...

5- 13 +3 - 5 = ...

6- لدى جودي 6 بنسات، أضاعت منها بنسين، فكم بقي معها:

أ. 2

ب. 8

ج. 10

د. 12

شكل (10-1): مسائل توضيحية - نوقشت في صدق المحتوى في اختبار رياضيات للصف الأول الأساسي.

تقييم ما إذا كانت خصائص الفقرات تطابق الأهداف أو محتوى المجال الخاص بها وثانية يجب أن يكون المحتوى، ومستوى التعقيد، وصيغة السؤال (المثير) وصيغة الاستجابة مناسب لهذه الاعتبارات.

4) كيف يجب تلخيص النتائج، أن السؤال عما إذا كانت مجموعة الفقرات تمثل النطاق المُعطى بشكل مناسب يعد قراراً نوعياً أكثر منه كمياً، وبغض النظر فإن المعاملات الكمية لتلخيص قرارات الحكام على الفقرات منفردةً قد تكون مفيدة، وتتضمن بعض المعاملات المقترحة لهذا الهدف ما يأتي:

أ- نسبة الفقرات المزاوجة للأهداف.

ب- نسبة الفقرات المزاوجة للأهداف بتقديرات عالية الأهمية.

ج- الارتباط بين الأوزان النسبية للأهداف والفقرات التي تقيس هذه الأهداف (Klein & Kosecoff, 1975).

د- معامل التوافق بين الفقرة والهدف (Rovinell & Hambeltion, 1977).

هـ- نسبة الأهداف التي لم تقيم بأي فقرة في الاختبار.

ولأن المعاملات المذكورة أعلاه تعتمد على مسوغات مختلفة جداً، فلا يوجد هنالك سببٌ لافتراض أنها تؤدي إلى الاستنتاج نفسه حول درجة المطابقة بين مجموعة فقرات الاختبار

ونطاق المحتوى. ويجب أن تؤخذ خصائص عديدة لهذه المعاملات بعين الاعتبار في الاختبار وعند تفسير النتائج. فمثلاً يتطلب كلا العاملين الأول والثاني عدد كبير من الفقرات (100 أو أكثر) للتفسير ذي المعنى. ويتأثر المعامل الثالث بتباين عدد الفقرات التي تقيم كل هدف وأوزانها النسبية. وتحديداً لو أعطيت الأهداف جميعها أوزان نسبية متساوية أو أنها قيست جميعاً بعدد من الفقرات متساوٍ، فإن قيمة معامل الارتباط هذا = صفر.

والمعامل الرابع الذي اقترح من قبل (Hambelton, 1980) هامبلتون و هامبلتون وروفنيلي (Hambelton & Revenelli, 1977) يمكن استخدامه لتقييم إلى أي مدى يكون درجة صدق محتوى فقرة معينة بالنسبة لمجموعة الأهداف. وتعتمد هذه الصيغة على افتراض أنه في الحالة المثالية يجب أن تزاوج الفقرة هدفاً واحداً فقط من مجموعة الأهداف، وتعكس طريقة جمع البيانات هذا الافتراض وذلك لأنه تم توجيه الخبراء لمزاوجة الفقرة لكل هدف، وأن يعطوا القيمة (1) فيما لو كانت هنالك مزاوجة والقيمة صفر إن لم يكن متأكداً من المزاوجة والقيمة (-1) فيما لو لم تتطابق الفقرة مع الهدف بوضوح. ويمكن حساب معامل توافق الفقرة مع الهدف K بالصيغة [*]

$$(\mu - \mu_k)\ \frac{N}{2n-1} = I_{ik} \qquad (1\text{-}10)\ldots\ldots\ldots$$

حيث ترمز μ إلى عدد الأهداف و μ_k إلى متوسط تقدير الأحكام على الفقرة μ للهدف K ، و μ إلى متوسط تقدير الأحكام على الفقرة I لجميع الأهداف. وأكبر قيمة محتملة لتوافق الفقرة - الهدف تساوي (1.00)، ويتم الحصول عليها فقط عندما تتطابق الفقرة مع هدف واحد ومن قبل المحكمين جميعاً. وفي حالة تطابق الفقرة الواحدة مع أكثر من هدف فإن معامل التوافق يكون أقل من (1). وفي الحالة المثالية يجب أن تكون قيم (I_{ik}) لكل فقرة في الاختبار عالية وللهدف التي صممت لقياسه ومنخفضة للأهداف الأخرى.

أخيراً، فإن نسبة الأهداف غير المغطاة من خلال مجموعة فقرات يعد معاملاً لمدى جودة تمثيل محتوى النطاق الكلي بوساطة مجموعة الفقرات هذه. لاحظ أنه في الحالة المتطرفة عندما تتطابق فقرات الاختبار جميعها مع واحد فقط من الأهداف المهمة فإن المعاملات الأخرى تكون عالية جداً، ولكن يكون هذا العامل منخفضاً لذا فإنه يقدم معلومات مختلفة أساساً. ويجب تفسير معامل التناسب هذا عندما يكون عدد الأهداف التي يقيسها الاختبار قليلاً.

(*) الصيغة المقدمة هنا هي صيغة مبسطة للصيغة التي قدمها هامبلتون (1980) وهامبلتون وروفنيلي (1977).

قضايا في تقييم صدق المحتوى:

إحدى المشكلات الرئيسة المرتبطة بصدق المحتوى هو إمكانية المزاوجة بين الفقرات والأهداف بشكل جيد، إلا أن الأهداف المكتوبة لا تمثل نطاق الأداء بشكل مناسب والذي يستهدف مستخدم الاختبار إجراء استدلالات عنه. وقد اقترح كرونباخ (Cronback, 1971) منهجية لعلاج هذه المشكلة تسمى مضاعفة تجربة البناء "Duplicate Constrnchion Experement" التي تتم باستخدام فريقين مستقلين من مطوري الاختبارات لإجراء تحديد المحتوى نفسه، وأسس تجريب الفقرات ومحك تفسير البيانات، ثم يطلب من كل فريق تطوير اختبار لنطاق محدد مثل "معرفة وضع أداة نقل رموز الحركة" وبعد تطوير الاختبارين وتطبيقها على المفحوصين نفسهم يحسب متوسط مربعات الفروق لدرجات الاختبارين. وباستخدام نظرية الدرجة الحقيقية التقليدي:

$$\sigma^2_{e2} + \sigma^2_{e1} = \frac{(X_2{}^2 - X_1{}^2)\,\Sigma}{N}$$

حيث X_1، X_2 هي الدرجات على صيغتي الاختبار المتوازيين، و $\sigma e^2{}_1$ و $\sigma e^2{}_2$ تباين الخطأ لصيغتي الاختبار (وإثبات هذا سيترك تمريناً للقارئ في نهاية الفصل) ويمكن حساب تباين الخطأ لصيغتي الاختبار بطرائق الاتساق الداخلي، لذا فإن النسبة تحسب باستخدام المعاينة الإحصائية الذي يعتمد على مفحوصين عددهم N:

$$\frac{\sigma^2_{e2} + \sigma^2_{e1}}{n\,/\,{}^2(x_2 - x_1)\Sigma}$$

وعند اقتراب هذه النسبة من (1) فإنها تشير إلى أن الاختبارين المبنيّين من قبل الفريقين المستقلين متشابهان (من خلال تباين الخطأ)، وكما هو الحال في الاختبار المجزأ بطريقة التجزئة النصفية والمعد من قبل فريق واحد، ولذلك يوجد مؤشر إلى أن الأهداف المقاسة بالاختبار ليست بالإدراك للنطاق الذي قدم من قبل فريق واحد من مطوري الاختبارات. ولكن لسوء الحظ نادراً ما تستخدم هذه الطريقة لأنها نوعاً ما غير عملية للمواقف التي نحتاج فيها إلى صيغة اختبارية واحدة. ومع ذلك فإنه من المنطق استخدم هذه الطريقة عندما نحتاج إلى عدد من الصيغ البديلة والتي سيستخدم فيها عدد من فرق كتبة الفقرات ومطوريها بشكل من الأشكال.

قضية ثانية في صدق المحتوى يتعلق بتحيز بعض الفقرات لعرق أو سلاسة أو جنس ومناسبته لأحكام صدق المحتوى، نقول بحدة أن الحكم على صدق المحتوى يتعلق بمدى ملائمة فقرات الاختبار للنطاق المحدد. ومن جهة إذا تم تحديد النطاق على أنه معرفة ألفاظ

الإنجليزية يبدو أن خصائص المفحوصين (بما تتضمنه من اللغة الأم للمتحدث أو تعرضهم لتعلم الانجليزية) غير أساسية (أو ليست بموضوع) لأحكام المطابقة بين فقرات الاختبار وخصائص النطاق، ومن جهة أخرى فيما لو تم وصف النطاق على أنه القدرة على حل المشكلات القصصية في الحواس وتم عرض المثير اللفظي باللغة الإنجليزية لأطفال متحدثين بالأسبانية. وهنا لا يمكن الوصول إلى استنتاج دقيق من خلال درجات الاختبار حول المهارات الحسابية للمفحوصين. وظهرت المشكلة في هذا الموقف الفعلي لأن خصائص النطاق غير ملائمة وأدت إلى نتائج غير ملائمة. ومن الواضح أن النطاق هو القدرة على حل مشكلات قصصية معروضة باللغة الإنجليزية باستخدام مهارات رياضية معينة، وتبقى درجات الاختبار تمتلك صدق محتوى لهذا النطاق الأكثر تخصيصاً. السؤال هل هو مجدياً محاولة وصف أداء مفحوصين معينين في ذلك النطاق. ومثل هذا السؤال يعد قضية تربوية لا سيكومترية، لذلك فإن القرارات الأكثر عمومية يمكن إجراءها باستخدام هذا الاختبار، ويجب أن يكون مطورو الاختبارات حساسين للخصائص الأساسية للمفحوصين الذين طلب منهم النجاح في الاختبار كمتطلب سابق، وتضمين هؤلاء عند وصف نطاق السلوكات المعاينة بالاختبار.

قضية أخرى هي ما إذا كانت بيانات الأداء على الاختبار والفقرات يجب أخذها بعين الاعتبار عند الحكم على صدق المحتوى، تتبنى كرونباخ من جهة (1971, P:455) موقف "لا شيء في منطق صدق المحتوى يتطلب أن يكون النطاق أو المجال متجانساً في المحتوى. فالفقرات ذات الارتباطات الداخلية الأكبر قد تمثل زيادة معاينة لمجموعة فرعية معينة من السلوكات الموجودة في النطاق". ومن جهة أخرى يبدو غير منطقي توقع أن الفقرات المطابقة للهدف نفسه تكون ارتباطاتها على الأقل متوسطة. ومع ذلك مثل هذه البيانات قد لا يكون دليلاً كافياً لصدق المحتوى (أو غيابه)، وهي بالتأكيد أدلة مهمة لملائمة النطاق (أو غيابه)، وهي بالتأكيد أدلة مهمة لملائمة النطاق المحدد وعملية صدق المحتوى. فإذا وجد أن الفقرات التي تم الحكم عليها أنها مطابقة للهدف نفسه غير مرتبطة عملياً، هنا تكون العملية التي تم من خلالها مطابقة الفقرات والأهداف، ونوعية الفقرات وملائمة صياغة الأهداف معرضة للمسائلة.

أخيراً، مع التقدم سنلاحظ مفاهيم عديدة متناقضة تعرض من وقت لآخر في أدبيات القياس والمشابهة ولكن غير مماثلة لصدق المحتوى، أحدها الصدق السطحي (Moiser, 1947) والذي يفسر عموماً على أنه المدى الذي تظهر فيه الفقرات تقيس بناء ذا معنى للمفحوصين النموذجيين، وفي بعض الحالات لاستبانات الشخصية فمن غير المرغوب فيه أن يكون للمفحوصين قدرة على استنتاج ما يتم قياسه بفقرات الاختبار، ولكن لاختبارات التحصيل المدرسي والتوظيف فإن الصدق السطحي قد يحفز المفحوصين لتقديم أفضل ما

لديهم عندما يتبين أن الاختيار يقيس بناءً معنوياً أو مهماً. ومفهومين آخرين حديثين في الأصل هما صدق المنهاج وصدق التدريس (انظر 1978 و McClung) .يشير الأول إلى مدى مناسبة الفقرات لأهداف منهاج معين كما وضعت في السابق، والأخير يشير إلى مدى تقديم المعلم لتدريس المحتوى الخاص والمهارات المقيسة بفقرات الاختبار.

وقد ناقش كل من يالو وبوبام (Yalow & Popham, 1983) ظهور هذه المفاهيم بالنسبة لصدق المحتوى والدور الذي تلعبه في قرارات المحكمة الحديثة في الولايات والمستخدمة لحقوق الولايات في الحصول على دبلوما عالياً اعتماداً على درجات التحصيل الاختبارية.

الصدق المرتبط بمحك:

يهدف مستخدم الاختبار في العديد من الحالات استخلاص نتائج من درجات الاختبار لفحص السلوك على محك أداء معين لا يمكن قياسه مباشرة بالاختبار. على سبيل المثال فقد يرغب المشرف على القبول في الكلية أن يستخلص ومن خلال درجات اختبار القبول للأداء الأكاديمي في مساق عملي في الكلية، أو قد يرغب أن يستخلص ومن خلال درجات الاختبار حجم المبيعات للذين طبق عليهم الاختبار وفي مواقف البيع التي يشغلونها، وقبل استخدام درجات الاختبار لاتخاذ قرارات حول القبول أو التشغيل يجب أن يحصل صانع القرار على أدلة تشير إلى أن هنالك علاقة بين درجات الاختبار والأداء على المحك، فهذا النوع من الأدلة يمكن الحصول عليه من دراسة الصدق المرتبط بمحك.

ويسير التصميم العام لدراسة الصدق المرتبط بمحك حسب الخطوات الآتية:

1) تحديد سلوك المحك المناسب وطريقة لقياسه.

2) تحديد عينة مناسبة ممثلة من المفحوصين الذين سيستخدم الاختبار لفئتهم.

3) تطبيق الاختبار والاحتفاظ بدرجة كل مفحوص.

4) عندما تكون بيانات المحك مناسبة احصل على قياس للأداء على المحك ولكل مفحوص.

5) تحديد قوة العلاقة بين درجات الاختبار والأداء على المحك.

ومن المألوف والمعتاد التمييز بين نوعين من الصدق المرتبط بمحك: الصدق التنبؤي والصدق التلازمي. فالصدق التنبؤي يشير إلى درجة تنبؤ الاختبار بالقياس المحكي الذي قد يجري عند نقطة معينة في المستقبل على سبيل المثال: قد ترتبط درجات اختبار الاستعداد المدرسي (SAT) بمتوسط درجة الطالب بالكلية قدرهُ (0.40)، وبالتالي، فإن درجات اختبار الاستعداد المدرسي لها درجة صدق تنبؤي بالنسبة لمتوسط درجة الطالب بالكلية(CGPA)

وهذا الدليل على الصدق التنبؤي هو الأساس في تعديل استخدام درجات اختبار الاستعداد المدرسي (SAT) في قرارات القبول. وقد يرغب مرشدو القبول في الكليات بالاعتراف بالتلاميذ الذين ينجحوا في الكلية، وأن (CGPA) يعد مقياساً عملياً للتحصيل الاكاديمي، وبسبب العلاقة المثبتة بين (SAT) و (CGPA) فإن استخدام درجات SAT في قرارات القبول يجب تعديلها جزئياً على الأقل لعمل استنتاجات حول أداء المفحوصين المستقبلي. والصدق التلازمي يشير إلى العلاقة بين درجات اختبار وقياس محكي طبق الاثنان في الوقت نفسه. على سبيل المثال إذا تقدم التدرب على الطيران لاختبار ورقة وقلم في المعرفة بالطيران، ومن ثم أجريت ملاحظة وتقدير أداؤه في الطيران الفعلي. والعلاقة الإيجابية يمكن أن تكون مؤشراً للصدق التنبؤي لاختبار الورقة والقلم، والارتباط الكبير لدرجة كافية قد يعدل استخدام الاختبار ليحل محل نظام ملاحظة أقل فعالية وأكثر كلفة في تقويم الطيران.

مشكلات تطبيقية للصدق المرتبط بمحك:

هنالك عدة مشكلات محتملة قد تؤثر في دراسة الصدق المرتبط بمحك، وأكثر مشكلة تواجهنا هي تحديد المحك المناسب، وعدم كفاية حجم العينة، والتلويث، وقيود المدى، وعدم ثبات درجات المتنبيء أو المحك، وسنناقش كل مشكلة منها باختصار في الجز التالي.

مشكلة المحك: أشار ثورندايك (Thorndike, 1949) إلى أن مقاييس المحك يمكن اعتبارها محكات فورية، أو متوسطة أو قطعية. ويبدو أن المحك الفوري متوافر بسهولة وسهل القياس نسبياً (مثل الدرجة في مساق معين، وتقدير المشرفين عن المرضين الذي يُعطوا جرعات تحت الجلد، والوقت اللازم لسكرتيرة لطباعة ثلاث رسائل قياسية وتغليفها). لسوء الحظ مثل هذه المحكات ليست تامة بدرجة كافية أو مهمة لاستخدامها كمحك منفرد لتصديق اختبار صمم للتنبؤ بالأداء على المجال الذي نهتم به. بالمقابل يتميز المحك القطعي على الأغلب بأنه ذو أهمية أساسية ولكنها صعبة للغاية في التحديد الإجرائي والقياس، وأمثلة على مثل هذا المحك قد يكون "الكفاءة الجراحية" أو "فعالية التدريس" أو "الاستقلالية في نشاطات الحياة اليومية"، وسيلحظ القارئ أن هذه المحكات عبارة عن بناءات فعلية يجب تحديدها إجرائياً بوساطة متغيرات يمكن قياسها بصورة مباشرة بدرجة أكبر.

على سبيل المثال أن أردنا أجراء تصديق اختبار للتنبؤ بفاعلية معلمي المستقبل الصفي، بشكل مثالي يجب قياس المحك المتأخر مع ملاحظات متكررة لأداء المعلمين في الصف لفترة سنوات ما بعد إجراء عملية التدريج، ولأن هذا المحك غير عملي بدرجة كبيرة لقياسه، سنواجه باستبداله بمحك متوسط من خلال تقدير طلبة المعلمين المقصودين خلال فترة دراستهم عند هؤلاء المعلمين. فالمحك المتوسط قد يكون الدرجة التي حصل عليها التلميذ على مشروع يتطلب أعداد خطة ودرس وتقديم محاضرة للرفاق في صفوف الدراسة الجامعية الأولية. مع ملاحظة

أنه كلما كان الحصول على قياس أسهل للمحك فإن أهميته أو تقريبه للمحك القطعي قد تخفض من الحاجة إلى المحك القطعي. وهذه تولد معضلة في التخطيط لدراسة الصدق المرتبط المحك فإن كان قياسه سهلاً نختار المحك الفوري، وقد لا يكون ارتباطه عالياً بالمحك القطعي الذي نهتم به، ومن جهة أخرى فعند التأشير لمحك قطعي فإن الجهد المطلوب لتحديد هذا المحك إجرائياً وكلفة الحصول على قياس ثابت له قد يكون مكلفاً وصعب جداً، لذا فإن اختيار المحك يجب أن يتسم بالحكمة وحسن التمييز، أي أنه:

(1) يمكن قياس ثباته ضمن حدود الوقت وكلفة الدراسة. (2) له علاقة بالمحك القطعي الذي يهتم به مستخدم الاختبار.

حجم العينة: عند حساب معاملات الصدق لعينات صغيرة، يكون خطأ المعاينة كبيراً نسبياً، وقد تنخفض القدرة الإحصائية الاستدلالية المستخدمة في تحليل بيانات دراسة الصدق بشكل أساسي. فمثلاً وجد كل من شميدت وهنتر ويوري (Schmidt,Hunter,urry,1976) انه لحجم عينة يتراوح بين (30–50) فإن المتنبيء الذي له مستوى صدق مقبول في المجتمع قد يكون له مستويات صدق مقبولة لعينات حجمها بين 25% الى35% ضعف هذا الحجم. وقد اقترح هؤلاء الحاجة إلى حجم عينة 200 أو أكثر ليعكس مستويات صدق بيانات المجتمع بدقة على الأقل بـ 90%، وفي السنوات الأخيرة توجد توقعات قانونية ومتخصصة عند مطوري الاختبارات ومستخدميها لإثبات وبرهنة صدق درجات اختبار في الاختبار الشخصي أو التربوي في كل مكان عمل خاص أو تربوي تستخدم فيه الاختبارات، لذلك وضمن الفترة الزمنية المعقولة لإجراء دراسة الصدق، قد يكون هنالك فريقاً من المستخدمين في المؤسسات الخاصة والمعاهد التربوية والذي يكون صغيراً إلى الحد الذي يسمح بتقدير مناسب لمعامل الصدق. وفي مثل هذه الحالات قد يكون التحقق من تعميم صدق المتنبيء بديل لإجراء دراسة الصدق بعينة غير مناسبة، وذلك بمراجعة دراسات الصدق المرتبط بمحك والمدونة في الدراسات لمتنبيء بعينة أو المتنبآت المشابهة والمحكات المشابهة للمحك المستخدم في الموقف الحالي.

التلوث بالمحك: يجب ألا تكون درجات الأفراد الذين قد يؤثروا على درجات المحك ضمن درجات المفحوصين على المتنبيء. فعلى سبيل المثل: فإن تعلم الأستاذ لأسلوب تصحيح اختبارات القبول والمثبت صدقها قد يؤثر على خبرات التدريس أو التطبيق، وهذا ما يعرف باسم تلوث المحك، من جهة فإن هذا التلوث قد يؤدي إلى زيادة في العلاقة بين درجات الاختبار ودرجات المحك (ارتفاع معامل الارتباط فيما بينهما)، وهذا الارتفاع يكون مصطنعاً فيما لو كان اتجاه الأستاذ لاعتبار الطلبة الذين حصلوا على درجات عالية في اختبار القبول

أنهم بارعون، وبالتالي يرى أعمالهم أفضل مما هي عليه، ومن جهة أخرى فإن تلوث المحك قد يقلل من العلاقة بين المتنبيء والمحك فيما لو علم الأستاذ أن طلبة معينين كانت درجاتهم منخفضة تسبباً على المحك، وبذل الأستاذ جهداً اكبر في تدريس هؤلاء التلاميذ، ولأنه من المستحيل التنبؤ بانتظام اثر التلوث فمن المهم اختزال فرص ظهوره.

تقييد المدى: لأن الصيغة الأكثر شيوعاً لمؤشر الصدق هو ارتباط درجات المتنبيء والمحك، فمن المهم تمييز أن تغيير مدى الدرجات على المحك أو المتنبيء قد يؤدي إلى تضعيف معامل الصدق الملاحظ، وهنالك طريقتان قد تؤدي تطبيقات الاختيار إلى تغيير التباين. الأول عند استخدام الاختيار الذي يجري تصديقه لأهداف الانتقاء قبل تأسيس عملية تصديقه (مع ملاحظة أن هذا مناقض لطرائق الصدق الموصى بها)، فجميع الأفراد الذين حصلوا على درجة أقل من نقطة معينة يتم رفضهم وبالتالي لن يتم استخدامهم في دراسة الصدق لأنه لا يتوافر لهؤلاء بيانات محكية، لذا فان حصر المدى الذي يظهر في المتنبيء قد يقيد التباين بسبب الاختيار الواضح على ذلك المتغير. كذلك فان التباين على المتنبيء قد يختصر (يختزل) فيما لو تم الاختيار بناءاً على أساس متغير آخر مرتبط بالاختبار المتنبيء. على سبيل المثال عندما يتم إجراء دراسة صدق لاختبار القبول في الكلية، ولكن خلال إجراء الدراسة تم قبول الطلبة بناء على أساس درجاتهم المدرسية والتي ترتبط بدرجاتهم على اختبار القبول، وهذا ما يعرف بالاختبار العرضي.

وقد يظهر تقييد المدى فيما لو كان القياس المتنبيء أو المحك سهلاً جداً لدرجة أن جميع المفحوصين حصلوا على درجات عالية (اثر السقف) أو انه صعب جداً لدرجة أن المفحوصين جميعهم حصلوا على درجات منخفضة (اثر الأرض)، وقد يكون تقييد المدى دالة لطريقة التصحيح المختارة(مثلاً عندما يعطي المرشدين علامات على تدريج مؤلف من أربعة نقاط، في حين أن تدريج بنقاط أكثر على المتصل يسمح بتمييز أكثر دقة، كذلك يعد التسرب الطبيعي مصدراً اخراً لتقييد المدى خاصة عندما يغادر الطلبة المتطرفين على التدريج (عند إحدى نهايتيه) المجموعة قبل جمع بيانات المحك على سبيل المثال: افترض أنه استخدم تدريج التطور السيكومتري للرضع في التنبؤ عن تطور لغة الأطفال (مرضى متلازمة دوانز) عند عمر سنتين، ويتوقع أن العديد من هؤلاء الأطفال أن يكون تطورهم اللغوي هو الأقل بسبب قسوة ظروفهم ليعيشوا إلى العمر الذي سيتم فيه الحصول على بيانات المحك.

وقد ناقش كل من ثورندايك (Thorndike, 1982,1949) وجو ليكسن (lord novick, Gulliksen,1968) طرائق حساب معاملات الصدق للمجموعة الكلية حتى عند الحصول على بيانات كاملة على المتنبيء في حين أن بيانات المتغير المحك متوافرة فقط لمجموعة فرعية

مختارة من المفحوصين بسبب الصعوبات التي يواجهونها في دراسة الصدق. وتتطلب هذه الطرائق افتراضات قد لا يمكن الدفاع عنها في المواقف العملية. وأحد هذه الافتراضات أن الانحدار الخطي ل y (المتغير المحك) على X (المتنبيء) هي نفسها عبر قيم X جميعها، وهو للمجموعات المختارة وغير المختارة. افتراض آخر هو أن $\sigma_{x.y}^2$ (تباين المحك للمفحوصين عند أي نقطة معطاة لـ X) هي نفسها لجميع قيم X. وقد لاحظ لورد ونوفيك (1968) أن الأخطاء في تصحيح معاملات الصدق قد لا تظهر فيما لو أهملت هذه الافتراضات خاصة عندما لا يكون الاختيار واضحاً على درجات المتنبيء. كخلاصة من الأفضل التخطيط لدراسة الصدق لتجنب تقليص أو اختصار التباين أن كان ممكناً بدلاً من الاعتماد على طرائق التصحيح الإحصائي في التغلب على هذه المشكلة

ثبات المتنبيء والمحك:

يبدو من المنطق في أي محاولة لتقييم درجة العلاقة بين المحك والمتنبيء أن خطأ القياس يجب أن يكون أقل ما يمكن. وفي الواقع يوجد علاقة مباشرة للارتباط بين Xو Yوثبات كل من المتنبيء X والمحك مثل:

$$\rho_{xy} = \sqrt{\rho_{xx'}}\sqrt{\rho_{yy'}} \qquad (2\text{-}10)...........$$

وإذا حصلنا على تقديرات ثبات المتنبيء و المحك وكانت $\rho_{xx'}$=0.81 و$\rho_{yy'}$ =0.64 ، يمكننا حساب أقصى ارتباط يمكن الحصول عليه بين هذه المتغيرات ويساوي 0 .72 وسيتم عرض لاشتقاق هذه العلاقة ومضامينها في تفسير نتائج دراسات الصدق. ويعد كافياً لغاية الآن ملاحظة أهمية جمع الملاحظات لكلا من المتنبيء والمحك وضمن ظروف تقلل أخطاء القياس لأقصى قدر ممكن، وكذلك فحص ثبات كل من المتنبيء والمحك. (وفي بعض الأمثلة النادرة سنلاحظ أن الارتباط بين X و Y يتجاوز $\sqrt{\rho_{xx'}}\sqrt{\rho_{yy'}}$ ويظهر هذا عندما يكون تقدير الثبات أقل من تقدير معامل الدقة لأحد أو كلا المتغيرين أو عندما لا تتحقق فرضية الارتباط بين الأخطاء على X وY).

ومع ذلك فمن المرغوب به الحصول على قياسات ذات ثبات عالية لكلا من الاختبار والمحك، وليس من الضروري صحة القول بأن القوة التنبؤية للاختبار تزداد بازدياد الاتساق الداخلي. ويمكن إيجاد مناقشة أوسع لهذا في لوفنجر **(Loevinger, 1954)** ولورد ونوفيك **(Lords Novick, 1968)** وايبل **(Ebd, 1968)** و هورن **(Horn, 1968)** وكين **(Kane, 1982)** باختصار نلاحظ أنه كلما كان الاختبار المتنبيء أكثر تجانساً في المحتوى (بزيادة الاتساق الداخلي) فإن الارتباط بين

الفقرات يزداد. وبشكل مثالي، حتى يعزى أكبر قدر من التباين إلى المتغير المحك، يجب أن يكون ارتباط الفقرات عالياً مع المحك ولكن ليس فيما بينها. (Jucker,1946) خذ مثلاً المثال المتطرف للاختبار المؤلف من فقرتين عندما يكون الارتباط بين الفقرتين = 1.0 ففي هذه الحالة لا تؤدي الفقرة الثانية إلى زيادة تباين المحك والذي لا يعزي إلى الفقرة الأولى. وهذا التناقض في زيادة الاتساق الداخلي في الوقت الذي ينخفض فيه الصدق ولدّ قضية ستؤخذ بعين الاعتبار في الفصل 14.

تدوين نتائج الصدق المرتبط بمحك وتفسيرها:

يمكن تحليل البيانات المتجمعة من دراسة الصدق المرتبط بمحك بعدة طرائق، فإن كان قياس المحك على توزيع متصل فإن معامل ارتباط بيرسون بين درجات الاختبار وقياس المحك يمكن حسابه، وهذا المعامل يطلق عليه اسم معامل الصدق. وفي بعض الحالات يمكن أن يكون المتغير المحك فئوي، ومثال ذلك عندما يكون المحك ثنائي الدرجات كأن يكون إكمال متطلبات النجاح لبرنامج الماجستير في علم النفس التحليلي أو الفشل في إكمال متطلبات النجاح للبرنامج نفسه.

وفي هذه الحالة فإن متوسط درجة المتنبئ للناجحين يمكن مقارنته بمتوسط درجات الذين لم ينجحوا، والفرق بين درجة اختبار القبول والنجاح أو الفشل في نهاية البرنامج (كبديل: الطرائق الارتباطية المستخدمة للمتغيرات المتصلة والمتقطعة التي وصفت في فصل 14). وعندما يتم تصنيف الأداء على كل من المتنبئ والمحك (مثل ناجح – راسب على المتنبئ، النجاح – الفشل على المحك) يمكن تدوين إحصائي مثل معامل فاي أو أي طريقة ارتباطية مناسبة للاستخدام مع البيانات الفئوية.

وقد لا يكون معامل الارتباط بين درجات الاختبار والمحك كافياً لمساعدة مطوري الاختبار في الحكم على مستخدميه درجات الاختبار لاحتياجاتهم الخاصة. ومن المعلومات الإضافية التي قد تدعم مستخدمي الاختبار في تفسير نتائج دراسة الصدق تتضمن معامل التحديد والخطأ المعياري للتقدير وجداول التوقع. وفي الفصل الثاني لاحظنا أن مربع معامل الارتباط يستخدم في تفسير أهمية العلاقة بين متغيرين اثنين. ويسمى معامل الارتباط بين درجة الاختبار ودرجة المحك معامل الصدق، ويسمى مربعه معامل التحديد، فإن كان الارتباط بين درجات الاختبار ومقياس الرضى الوظيفي = 0.60، فإن معامل التحديد = 0.36 ويؤشر إلى أن 36% من التباين في الأداء الوظيفي يُعزى إلى تباين الأداء على الاختبار المتنبئ.

وعندما تستخدم درجات الاختبار للتنبؤ بالأداء المحكي للأفراد فمن المهم تدوين الخطأ

جداول التوقع:
يوجد اختبار تنبؤي (اختبار القبول)
استخرج من الجدول عدد الطلاب الذين سينجحون في أصل ...
وفقاً للمعدل التراكمي لعدة سنوات

المعياري للتقدير والمناقش في فصل 2 تذكر أن معادلة التنبؤ بدرجة الفرد y على المتغير المحك هي:

$$\hat{\mu}_y + (\hat{\mu}_x - x)\,(\hat{\sigma}_x / \hat{\sigma}_y)\,\hat{P}_{xy} = y^1$$

حيث أن $\hat{P}_{xy}$ ترمز إلى معامل الصدق، و $\hat{\sigma}_y$ ترمز للانحراف المعياري على درجات المحك، $\hat{\mu}_y$ ترمز إلى متوسط درجة المحك و $\hat{\mu}_x$ و $\hat{\sigma}_x$ يرمزان إلى المتوسط والانحراف المعياري على المتنبئ المحسوب من عينة دراسة الصدق. ولإيجاد فترة الثقة حول الدرجة المتنبأ بها للفرد y^1سنستخدم الخطأ المعياري للتقدير في عينة الصدق:

$$\sqrt{\hat{P}^2_{xy} - 1}\;\hat{\sigma}_x = \hat{\sigma}_{yx}$$

ولمستخدم الاختبار الذي حدد قيم y^1 المحسوبة للمفحوص يمكن أن تكون بدرجة ثقة 68% عندما تكون درجة y الفعلية تقع في الفترة التقريبية $\hat{\sigma}_{yx} + y^1$ ، وبـ 95% ثقة عندما تقع درجة y الفعلية في الفترة التقريبية $\sigma_{yx}\ 2 + y^1$ (ومثال للحساب والتفسير للخطأ المعياري للتقدير مبين في فصل 2، انظر معادلة 2 – 20).

أخيراً، في تفسير مدى دقة درجات الاختبار للمفحوصين ضمن مدى درجات معين في الاختبار أو المحك، فقد يغير أحياناً استخدام جدول التوقع والذي يتضمن احتمالات درجات المحك المختلفة للمفحوص مع درجة معينة على الاختبار المتنبئ. ومثل جداول التوقع هذه مبين في جدول (10-1).

وفي هذا المثال استخدمت استبانة مهارات المهنة للتنبؤ بالأداء الوظيفي للطلبة العاملين في برنامج تدريب القوة البشرية والعدد في كل خلية يدل على نسبة المتدربين لكل مستوى درجة في استبانة مهارات المهنة والذين أرسلت إليهم توصيات المشرف (المحك) مبين في رأس الجدول. ويمكن تفسير القيم المبينة في الجدول على النحو الآتي:

جدول (10-1): جدول توقع يبين نسبة المتدربين في كل فئة درجات عند مستويات مختلفة:

درجة مهارات المهنة	استخدام بمستوى أعلى من الدخول	استخدام عند مستوى الدخول	استخدم في الامتحان الصارم	غير مستخدم	عدد المتدربين
110-101	-	-	-	100	3
120-111	-	-	-	100	5
130-121	-	-	33.3	66.7	6
140-131	-	-	30.0	70.0	20
150-141	-	5.6	22.2	72.2	18
160-151	3.8	11.5	34.6	50	26
170-161	8.6	28.6	28.6	34.2	35
180-171	8.5	34.0	27.7	29.8	47
190-181	22.5	47.5	23.8	6.2	80
200-191	53.6	40.2	6.2	-	97
210-201	72.3	21.4	4.5	1.8	112
+211	100	-	-	-	26

المتدرب الذي تقع درجته في المدى151–160 على استبانة مهارات المهنة فرصته 3.8% لا يكون مستواه مستخدماً أعلى من مستوى الدخول وفرصته 5.1% لأن يكون مستخدماً عند أدنى مستوى للدخول وفرصة 34.6% لأن يكون مستخدماً في الامتحان الصارم، وفرصة 50% لان يكون غير مستخدم.

ولمعظم الاستخدامات التحليلية والتربوية والشخصية فان درجات اختبار واحد تكون بياناتها غير كافية لاتخاذ قرارات مهمة حول المفحوصين المنفردين.وفي حالة تحديد واحد أو أكثر من المتنبآت المفيدة، فان سؤال الصدق يكون، هو ما إذا أمكن للمتنبيء الجديد أن يؤدي إلى تطوير دال في تحسين التنبؤ بالأداء المحكي عند أضافته للمتنبآت الأخرى. والطرائق الإحصائية لتناول هذا السؤال مطروح في فصل 11. ويناقش الفصل 12 اعتبارات أخرى مهمة في تفسير معلومات الصدق والذي يتعلق بقرارات الاختيار.

صدق البناء:

نوقش في الفصل الأول استخدام الأبنية النفسية كأساس في تطوير الاختبارات. وحدد البناء النفسي على انه نتاج للتخيل العلمي، وفكرة طورت لتسمح بالتصنيف والوصف لبعض السلوكات الملاحظة بشكل مباشر، ولا يمكن ملاحظة الأبنية النفسية مباشرة. ومن الأمثلة على الأبنية النفسية الذكاء، والإبداع، والاستقلال عن المجال، والانبساط -الانطواء. وقد لاحظ لورد ونوفيك (1968) انه من الأهمية بمكان تحديد البناء في مستويين: الأول يجب تعريفة اجرائياً (أو تحديده لغوياً)، ويكون عادة بتخصيص الطرائق المستخدمة في قياس البناء. وأمثلة على الاختبارات مثل اختبار ستانفورد-بينييه، واختبار عزل الترابطات، واختبار تفهم الموضوع، وجميع هذه توضح التعريفات الإجرائية لأبنية مسماة، بالإضافة إلى أن البناء النفسي يتطلب تعريفاً يوضح ترتيب كلمات الجملة في أشكالها وعلاقاتها الصحيحة وذلك باقتراح علاقات بين مقاييس البناء 1) مع مقاييس أبنية أخرى في النظام النظري و 2) مقاييس للمحكات الخاصة في العالم الحقيقي. بكلمات أخرى، فإن التعريف الإجرائي ليس كافياً، ويجب تبيان المعنى أو الأهمية بوضوح من خلال وصف كيفية ارتباطها بالمتغيرات الأخرى.

مع ذلك فانه تم تجميع أدلة صدق البناء من خلال سلسلة دراسات، وتتضمن هذه العملية الخطوات الآتية:-

1- صياغة فرضية أو أكثر تبين الاختلافات المتوقعة في الخصائص الديموغرافية أو محكات الأداء، أو مقاييس الأبنية الأخرى ذات العلاقة بمحك أداء تم تصديقه وذلك لمن هم مختلفين في البناء المراد إجراء دراسة صدق له. وهذه الفرضيات يجب أن تعتمد على نظرية صياغتها واضحة ويقع البناء ضمنها، وتزودنا بتعريف خصوصي للبناء.

2- اختر أو طور أداة قياس تتألف من فقرات تمثل سلوكات مخصصة وواضحة حسياً للبناء.

3- اجمع بيانات تجريبية تسمح باختبار العلاقة الافتراضية.

4- حدد ما إذا كانت البيانات مطابقة للفرضية مع الأخذ بعين الاعتبار مدى إمكانية تفسير النتاجات الملاحظة بوساطة النظريات المنافسة أو التفسيرات البديلة (واحذفها إن أمكن).

ويجب أن يكون واضحاً من القائمة السابقة أن صدق درجات الاختبار وصدق النظرية حول طبيعة البناء الذي نهتم به لا يمكن الفصل بينهما، وإن كانت العلاقة المفترضة مثبتة كما تنبأت بها النظرية فإن كلاً من البناء والاختبار الذي يقيسه مفيداً، وإن لم نتمكن من إثبات

الفرضية بوساطة دراسة الصدق فإن مطور الاختبار لا يستطيع معرفة ما إذا كان هنالك قصوراً في البناء النظري أو الاختبار الذي يقيسه أو كلاهما. (في هذه الحالة صدق البناء يشبه نوعاً ما وضع رهان على كلاً من النظرية والاختبار الذي يقيس البناء والذي يجب بناءوه جيداً قبل أن يتم الدفع من قبل السيكومتري المغامر.

طرائق للصدق البنائي:

يتطلب صدق البناء تجميع وتصنيف أنواع عديدة من الأدلة أكثر من أي أسلوب صدق أكثر آخر. وهنالك أربعة طرائق واسعة الانتشار لصدق البناء ستناقش هذا.

الارتباط بين مقياس للبناء والمقياس المصمم:

مثال تقليدي هو محاولة تأسيس أدلة ارتباطية للعلاقة بين درجات اختبارات الذكاء ومقاييس الأداء المدرسي أو الوظيفي. ومع أنه يبدو غير منطقي الموافقة على أن الذكاء والتحصيل المدرسي تعد بناءات متماثلة، إلا أنه يمكن الموافقة على (أو يجب) أن يكون هنالك علاقة فيما بينهما. وإن لم يكن الموقف هكذا فإن أهمية الذكاء كبناء تتلاشى، ولا يوجد مؤشرات عامة تميز ما يتألف منه الدليل المناسب لصدق البناء في الدراسات الارتباطية. أما الارتباطات الفردية فيمكن بالطبع اختبار دلالاتها الإحصائية، ويمكن تدوين نسبة التباين المشترك وهكذا..، ومثل هذه المعلومات لوحدها ليست كافية دون المقارنة مع مدى القيم المدونة سابقاً من قبل الآخرين الذين طوروا مقاييس للبناء نفسه أو بناءات مشابهة له. ومثل هذه المعلومات تساعد مستخدمي الاختبارات في تقييم قوة الأدلة المقدمة لصدق بناء الاختبار. وفي العديد من الحالات يستخدم الأسلوب الارتباطي تطبيق الانحدار المتعدد، لذا يمكن تقييم مساهمة البناء الهدف في التباين على المحك نسبة إلى مساهمة المتغيرات الأخرى (وهذه ستناقش في الفصل 11). كذلك اقترح دارلنغتون (Darlingtone, 1970) طريقة يفترض فيها مطوّر الاختبار قيماً نسبية للارتباطات بين مقياس البناء والمتغيرات الأخرى الملاحظة من أجل تأسيس أدلة حول صدق البناء.

التمييز بين المجموعات:

مثال على هذا الأسلوب مقارنة متوسط درجات الذكور والإناث على مقياس التقدير الذاتي لإدراك دور الجنس لرؤية ما إذا كانت تختلف بالاتجاه المبين بالفرضية والفشل في إيجاد فروق متوقعة قد يظهر شكوك حول البناء لإدراك دور الجنس أو ملائمة هذه الأداة كقياس للبناء أو كلاهما. مثل هذه الدراسات عبارة عن دراسات تجريبية من حيث التصميم، وتهدف

التطبيق على الأفراد الذين تم إخضاعهم لمعالجة معينة صممت لتبديل موقفهم على البناء عن أولئك الذين لم يتعرضوا للمعالجة. فإن لم توجد فروقات في مثل هذه الدراسات فإن التفسيرات المحتملة هي فشل في النظرية، و/ أو البناء، و/ أو عدم ملائمة الأداة في قياس البناء و/ أو فشل في المعالجة.

التحليل العاملي:

يتطلب هذا الأسلوب بشكل عام الحصول على مجموعة قياسات عددها n على المفحوصين نفسهم، وحساب مصفوفة ارتباطات n بين هذه القياسات، ومن ثم استخدام تقنيات التحليل العاملي لتحديد أقل عدد من المتغيرات (والتي تسمى عوامل) يسبب التباين في مجموعة المتغيرات N الأصلية. ويعرض الفصل 13 تفصيل أكثر للتحليل العاملي، ولكن هنالك تطبيقين عامين في صدق البناء ستذكر باختصار هنا. في الحالة الأولى يتم معاملة مصفوفة الارتباطات الداخلية للفقرات (عدد فقرات الأداة = n) لتحديد ما إذا كانت الاستجابات على الفقرات تتجمع معاً في نمط معين يمكن التنبؤ به أو منطقي في ضوء التركيب النظري للبناء الذي نهتم به. ويشكل تباين الاستجابات على الفقرات تجمع يمكن أن يُعزى إلى التباين عبر المفحوصين على عامل مألوف مسبب للتباين. ومثل هذا العامل الذي لا يمكن ملاحظته مباشرة يمكن اعتباره بناء مقترح من خلال مجموعة معينة من الملاحظات التجريبية. والقضية هي ما إذا كانت الأبنية المحددة تجريبياً من خلال التحليل العاملي مناظر للأبنية النظرية التي افترضها مطور الاختبار أثناء تطوير اختباره. فعلى سبيل المثال، في اختبار تحصيلي في الرياضيات[3] إذا كانت الفقرات تتطلب تنفيذ حسابات رياضية بسيطة تجمعت في عامل واحد، وفقرات تتطلب حل مسائل أكثر تعقيداً لمشكلات بكلمات تجمعت في عامل آخر، فإن هذا قد يثبت توقع مطور الاختبار. وإذا تجمعت عدة فقرات وظهر أنها تقيس الشيء نفسه (مثل حقائق ضرب عدد من منزلة واحدة) في عامل آخر فإن هذا يبرز أسئلة حول صدق البناء المقاس لهذا التجمع من الفقرات. وفي الحالة الثانية يمكن معاملة مصفوفة ارتباطات مجموعة اختبارات أو قياسات مختلفة عددها n لتحديد مدى الارتباط عبر الدرجات الملاحظة في هذه المقاييس لأن يُعزى لتباين عامل مشترك واحد أو أكثر. ويوضح الفصل 13 مثال مفصل لهذا النوع من صدق البناء من خلال التحليل العاملي. وثانية، القضية الحاسمة هي

---

(3) قدم هذا المثال على أنه توضيح مبسط، ولكن يجب ملاحظة أن التحليل العاملي لاستجابات الفقرات الثنائية تظهر مشكلات نوقشت في الفصلين 13، 14.

ما إذا كانت الاختبارات الفرعية أو الاختبارات التي يفترض أنها تقيس البناء نفسه تم تحديدها تجريبياً على أنها تقيس عاملاً مشتركاً.

مصفوفة السمات الطرق المتعددة:

طور كامبل وفيسك (Campell & Fiske, 1959) هذا الأسلوب على أنه يتعلق بملائمة الاختبارات كمقاييس للبناء لا ملائمة البناء كما هو محدد من خلال البراهين النظرية المتنبأ بها والمرتبطة بمقاييس بناءات أخرى. وفي هذه التقنية يجب على الباحث أن يفكر في طريقتين أو أكثر لقياس البناء الذي يهتم به، وعلاوة على ذلك يطلب منه تحديد بناءات أخرى مختلفة تماماً يمكن قياسها بشكل مناسب بالطرائق نفسها المطبقة على البناء الذي يهتم به.وباستخدام عينة واحدة من الأفراد يتم الحصول على قياسات لكل بناء بكل طريقة، ثم تحسب الارتباطات بين كل زوج من القياسات، وكل معامل ارتباط يعرف على أنه أحد الأنواع الثلاثة الآتية:

1. **معاملات الثبات**: هي معاملات ارتباط بين قياسات البناء نفسه باستخدام طريقة القياس نفسها، ويجب أن تكون عالية.
2. **معاملات الصدق التجميعي**: هي معاملات ارتباط قياسات البناء نفسه باستخدام طرائق قياس مختلفة، وهذه يجب أن تكون عالية، ولكن احتمالية الضعف الناجمة عن عدم ثبات طرائق القياس يجب أخذها بعين الاعتبار.
3. **معاملات الصدق التمييزي**: هي معاملات ارتباط بين مقاييس بناءات مختلفة باستخدام طريقة القياس نفسها، ويطلق عليها اسم معاملات الطريقة الواحدة – السمات غير المتجانسة وهذه يجب أن تكون أقل بصورة أساسية من كلا معاملات الثبات أو معاملات الصدق التجميعي.

ولتسهيل المقارنة عبر الأنواع المختلفة من المعاملات يتم ترتيب المعاملات في مصفوفة السمات – الطرائق المتعددة مثل المبينة في جدول (10–2). والبيانات المعروضة في هذا الجدول تم تدوينها في دراسة موشير(Mosher, 1968) وتتعلق بثلاثة بناءات مختلفة أسماها الشعور بإثم الجنس Sexguilt والعدائية Hostility Guilt والشعور الخلقي Morality Conscience، وقيست هذه بعدة طرائق: اختبارات صح – خطأ والاختيار الإجباري، والجمل غير المكتملة، وطبقت على عينة مؤلفة من 62 مفحوصاً جميعهم من الإناث، ووضعت معاملات الثبات بين قوسين وتشكل قطر مربع مصفوفة الارتباط، ووضع خط تحت معاملات الصدق التجميعي لأن هذه جميعها عبارة عن ارتباطات بين قياسات تستخدم الطريقة نفسها ويوجد في السطور (2 أ إلى 2 ج) و (3 أ إلى 3 ج) معامل صدق تجميعي واحد في كل صف. ففي

صف 2: أ يمثل المعامل (0.86) الارتباط بين قياس الشعور باثم الجنس المقاس بطريقتي الصح – الخطأ والاختيار الإجباري.

وفي كل صف من الصفوف الستة الدنيا يتبين أن معاملات الصدق التجميعي أكبر من معاملات الصدق التمييزي، ويتوافر الأدلة التي تدعم النظرية المقترحة لأنواع الإثم المختلفة. إضافة إلى أن القيم النسبية للارتباطات عبر السمات المختلفة تقع في النمط نفسه في 7 من 9 لمثلثات السمات غير المتجانسة. (وقد اقترح كل من كامبل وفيسك بصورة نموذجية نمط العلاقة نفسه عبر السمات يجب ملاحظته عبر المثلثات غير المتجانسة جميعها). ومع أن كامبل وفيسك أوصيا بالمعاينة أو الفحص البصري لتقييم بيانات صدق البناء في مثل هذه المصفوفة، والتي قد تكون مشكلة بسبب خطأ المعاينة. وحديثاً، اقترحت طرائق تحليلية إضافية قد تؤدي إلى تفسيرات أوضح لمثل بيانات هذه المصفوفة، وذلك لعينات أكبر بشكل ملحوظ. (انظر على سبيل المثال: (Lomax & Algina, 1979)و (Marsh and Hoce-var,1983)و (Schmitt ,1978).

نوع أخر من أدلة صدق البناء هو ما إذا كانت الملاحظات للفرد على البناء لا تتباين عبر طرائق القياس المختلفة. فقد اقترح كين (Kane, 1982) التحقق من هذه النقطة من خلال عناصر

جدول (10 –2):جدول توضيحي لبيانات مصفوفة السمات – الطرائق المتعددة من دراسة موشير(4).

	طريقة 1 السمة أ	طريقة 1 ب	طريقة 1 جـ	طريقة 2 السمة أ	طريقة 2 ب	طريقة 2 جـ	طريقة 3 السمة أ	طريقة 3 ب	طريقة 3 جـ
1– صح – خطأ									
أ– الشعور بالإثم الجنسي	(0.95)								
ب– العدائية	0.28	(0.86)							
ج– الشعور الخلقي	0.58	0.39	(0.92)						
2– الاختيار الإجباري									
أ– الشعور بإثم الجنس	0.86	0.32	0.57	(0.95)					
ب– العدائية	0.30	0.90	0.40	0.39	(0.76)				
ج– الشعور الخلقي	0.52	0.31	86.0	0.55	0.26	(0.84)			
3– الجمل غير المكتملة									
أ– الشعور بإثم الجنس	0.73	0.10	0.43	0.64	0.17	0.37	(0.48)		
ب– العدائية	0.10	0.63	0.17	0.22	0.67	0.19	0.15	(0.41)	
ج– الشعور الخلقي	35.0	0.16	0.52	0.31	0.17	0.56	0.41	0.30	(0.58)

62 = N

عن د. ل. موشير. قياس الإثم باستبانات التقرير الذاتي J. of consulting & clinical .أعيد طباعتها 1968 من قبل رابطة السيكولوجيين.

تحليل التباين الناتجة عن تطبيق نظرية إمكانية التعميم (عرضت في الفصل 8). ويفترض أن كل قياس يؤخذ على الفرد يمثل عينة عشوائية من نطاق القياسات المحتملة التي يمكن الحصول عليها. وهذا النطاق قد يتغير عبر أبعاد متعددة. تسمى عوامل، والتي يتم تشكيلها بطرائق قياس مختلفة. ويعتمد تحديد الأبعاد المناسبة في صدق البناء على النظرية الخاصة بالباحث للبناء والنطاق الذي سيتم التعميم عليه. دعنا نوضح ذلك من خلال تقييم بناء ما مثل مهارات الاستيعاب القرائي للمفحوصين، فمن الممكن قياس هذا البناء باستخدام صيغ فقرات متنوعة (اختيار من متعددة، صح – خطأ، نهاية مفتوحة، المزاوجة، ... الخ). وسنشعر بأن البناء أكثر أهمية أو قابلية للتعميم فيما لو كانت درجة المفحوص النسبية هي نفسها بغض النظر عن صيغ الفقرات المستخدمة في القياس. وإن رغبنا بمعرفة مدى جودة تعميماتنا عبر صيغ الفقرات سنصمم دراسة G (عرضت في الفصل 8) باستخدام صيغ فقرات مختلفة. افترض أننا صممنا هذه الدراسة وتم اختبار كل مفحوص بصيغ فقرات عددها N (في هذه الحالة نود القول بأن البيانات تم مطابقتها أو معايرتها على بُعد صيغة الفقرة). وإذا أردنا اعتبار صيغ الفقرات في دراسة G عينة عشوائية لصيغ الفقرات المحتملة جميعها والممكن استخدامها مستقلاً في دراسة D، وهنا تعد صيغ الفقرات بعد عشوائي، ومعامل إمكانية التعميم المناسب هو:

$$\frac{\sigma_p^2}{\sigma_e^2 + \sigma_{pf}^2 + \sigma_p^2} = P^2$$

حيث أن σ_p^2 ترمز لتباين الأفراد

و σ_{pf}^2 يرمز لتباين وتفاعل الأفراد والصيغة.

و σ_e^2 ترمز إلى عناصر التباين غير المفسر للدرجات.

ويشير كين (Kane, 1982) إلى أن هذا المعامل هو معامل صدق مع ملاحظة أنه يمكن تفسيره على أنه متوسط معامل الصدق التجميعي الناتج عن الاختيار العشوائي لطرائق مختلفة تقيس السمة نفسها من نطاق الطرائق الممكنة.

وقد يلاحظ القارئ أن تركيب معامل إمكانية التعميم هذا تقديراً لمعامل الصدق ولا يختلف في التركيب عن معامل إمكانية التعميم المقدم في فصل (8) على أنه تقدير للثبات. فهل هذا يشير إلى أن نموذج كين لا يميز بين الصدق والثبات؟ هنالك ثلاث طرائق مختلفة على الأقل للإجابة

عن هذا السؤال. الأول حدد كين بالتخصيص $(\sigma_e^2 + \sigma_{Pp}^2 + \sigma^2{}_p/(\sigma_{Pf}^2 + \sigma^2{}_p) = \sigma_e^2)$ على أنه معامل لتقدير الثبات، لذلك فإنه قيد تقدير الثبات لمعاملات إمكانية التعميم التي تكوّن طرائق القياس المختلفة بُعداً ثابتاً وليس عشوائياً (ويعد هذا تقديراً لما أطلق عليه لورد ونوفيك (Lord & Novick, 1968) معامل ثبات خاص). الثاني: هو أن التمييز بين معامل إمكانية التعميم على أنه تقدير للصدق أو الثبات هو مفاهيمي وسيعتمد على البناء ونطاق التعميم. كمثال: يمكن لمطور الاختبار أن يعتبر صيغ الاختبار المختلفة باستخدام الطريقة نفسها دراسة ثبات ولكن دراسة طرائق قياس مختلفة على أنها دراسة صدق.

التداخل عبر الأساليب المتعلقة بدراسة الصدق:

بعد استيعاب المادة المطروحة للأساليب الثلاثة المستخدمة للوصول إلى صدق الاختبار، فقد يبدو أن هذه الطرائق مقتصرة في تطبيقاتها على أنواع معينة من الاختبارات، فللوصول إلى استنتاجات غير متحيزة من خلال درجات الاختبار من الشائع استخدام أنواع عدة من مؤشرات الصدق. على سبيل المثال، افترض أن باحث يعتقد أن النجاح في مساقات العلوم الجامعية تعتمد على قدرة الطلبة على قراءة المادة التقنية مع فهمها. ويعتقد أيضاً أن الاختبارات التقليدية للاستيعاب القرائي غير مناسبة لتقييم هذه القدرة، وقد طور هذا الباحث اختبار "قدرة قراءة المادة التقنية" تتألف من نصوص مختارة من الكتب المقررة في الفيزياء والبيولوجيا في مستوى الكلية. واتبع كل نص بسلسلة أسئلة من نوع الاختيار من متعدد عن المعلومات المقدمة في النص. تعد هنا دراسة صدق المحتوى مناسبة لتحديد درجة ملائمة الأسئلة للمادة الموجودة في النصوص، وللتأكد من مدى تمثيل هذه النصوص للكتب المقررة في مساقات العلوم. ويمكن إجراء دراسة صدق مرتبط بمحك للتعرف على ما إذا كان الأداء على الاختبار مرتبط بالدرجات التي يحصل عليها الدارسون الجدد في مساقات العلوم. لذا فعند هذه النقطة يمكن استنتاج أن الدرجات العالية في الاختبار تعني قدرة قرائية تقنية عالية ولا يجب تعديلها. من جهة أخرى فإنه تم ملاحظة فرضية مهمة وبالتحديد بعض فقرات الاختبار تقيس معلومات علمية عامة يمكن أن يكون الطلبة قد اكتسبوها في العلوم التي درسوها في المرحلة الثانوية. ومثل هؤلاء التلاميذ تكون لديهم قدرة عالية على الإجابة عن الأسئلة بشكل صحي ودون قراءة النصوص على الإطلاق، وذلك عائداً إلى أعدادهم الجيد في المدرسة، وقد يحصل هؤلاء على درجات عالية في مساقات الكلية في العلوم، لذا فإن الصدق يكون أكثر اكتمالاً عندما يتضمن أدلة بأن الاختبار يقيس شيئاً مميزاً واضحاً من المعرفة العلمية العامة، ومن خلال الاستيعاب القرائي العام أو الاستعداد الأكاديمي، ومثل هذا الجهد في الصدق يجب أن يتضمن كل من صدق المحتوى والصدق المرتبط بمحك وصدق البناء.

إن إدراك الحاجة إلى أنواع عدة من الصدق وحقيقة أن البناءات يُستشهد بها في وصف نطاقات الأداء أو تعريف المحك قادت بعض المؤلفين مثل ميسك (Messick, 1981) ليأخذ موقفاً بأن "صدق البناء قد لا يكون كل الصدق ولكنه بالتأكيد قلب الصدق. والمناظرة حول أي نوع من المؤشرات هو الأكثر أهمية وأساسية نوقشت في أدبيات القياس لمدة نصف قرن تقريباً (انظر كمثال: Guion, 1977, 1978, 1980, Loevinger, 1947, and Tenopt, 1977).والموقف الذي نتبناه هنا هو أن نوع الصدق الأكثر أهمية يعتمد على النتائج المراد استخلاصها من درجات الاختبار.

والمناقشة حول أي نوع من مؤشرات الصدق هو الأكثر أساسية لتفسير درجات الاختبار قد يخدم بالفعل لحل الغموض في هذه النقطة. فيجب على مطوري الاختبارات ومستخدميها أن يفكروا من خلال نوع الأدلة الأكثر مستخدمية لدعم الاستنتاجات المراد استخلاصها من درجات الاختبار بشكل أفضل. وبعدها يجب تخطيط دراسات الصدق التي توفر مثل هذه الأدلة. والأكثر تفضيلا تضمين هذه الدراسات أنواع الصدق الثلاثة التي نوقشت في هذا الفصل.

معاملات الصدق للدرجات الحقيقية:

تفترض نظرية القياس التقليدية أن الأخطاء العشوائية للقياس غير مرتبطة، فإن كان هذا الافتراض صحيحاً فإن معامل الصدق بين الدرجات الملاحظة على الاختبار وقياس أخر نهتم به يكون أقل مما يجب إذا كان هنالك ارتباطاً بين الدرجات الحقيقية على القياسين. ويبين البرهان البسيط العلاقة بين معامل صدق الدرجات الملاحظة ومعامل صدق الدرجات الحقيقية المناظر له على المقياس نفسه. تذكر أنه في فصل 6 وضحنا بأنه يمكن التعبير عن الارتباط بين درجات الاختبار في موقفين اختباريين على أنه نسبة التباين المشترك للدرجات الحقيقية إلى حاصل ضرب انحرافاتها المعيارية بالاستدلال نفسه (أخطاء القياس غير مرتبطة). ويمكننا الآن التعبير عن الارتباط بين درجات الاختبار (X)ومتغير آخر Y بـ:

$$P_{xy} = \frac{\sum t_x t_y}{N \sigma_x \sigma_y} \qquad (3\text{-}10)\ldots\ldots\ldots$$

وإذا ضربنا هذا الكسر بـ $\frac{\sigma_{tx}}{\sigma_{tx}}$ و $\frac{\sigma_{ty}}{\sigma_{ty}}$

فإن القيمة لا تتغير وتصبح:

$$\rho_{xy} = \frac{(\sigma_{ty}) \quad \sigma_{tx} \quad \Sigma(t_y\, t_x)}{\sigma_{ty} \quad \sigma_x \quad \sigma_y\, \sigma_x\, N} \qquad \text{(4-10)..........}$$

ويمكن إعادة كتابة معادلة (10–4) كحاصل ضرب ثلاثة كسور منفصلة:

$$\rho_{xy} = \left(\frac{\sigma_{ty}}{\sigma_y}\right)\left(\frac{\sigma_{tx}}{\sigma_x}\right)\left(\frac{\Sigma(t_y\, t_x)}{\sigma_{ty}\, \sigma_{tx}\, N}\right)$$

ومن خلال تعريفات كل من معامل الارتباط ومعامل الثبات نصل إلى:

$$\rho_{xy} = \rho_{t_x t_y} \sqrt{\rho_{xx'}} \sqrt{\rho_{yy'}} \qquad \text{(5-10)..........}$$

ولأن قيم كل من $\rho_{xx'}$ و $\rho_{yy'}$ تكون أقل من 1 دائماً فإن معامل ارتباط الدرجات الملاحظة ρ_{xy} سيكون أقل من معامل الارتباط بين الدرجات الحقيقية بحل المعادلة (10–15) للحد $\rho_{t_x t_y}$ على النحو الآتي:

$$\rho_{t_x t_y} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}}\sqrt{\rho_{yy'}}} \qquad \text{(6-10)..........}$$

ويطلق على المعادلة (10– 6) أحياناً التصحيح من أثر التلويث (attenuation) لانها تؤدي إلى معامل صدق مصحح من أخطاء القياس على متغيري المتنبئ والمحك. وقد يكون أحياناً نوع من الفهم الخاطئ حساب هذا المعامل وتدوينه لدراسة الصدق المرتبط بمحك وذلك لأنه يتضمن درجة أكبر من الصدق الذي يمكن الحصول عليه بوساطة الاختبار المتوفر ومقياس المحك. ومع ذلك فهنالك موقف يكون معامل الصدق مضللاً ويجب تعديله. افترض أن باحث في دراسة صدق اهتم بمعرفة أي من المتغيرين المختلفين أكثر ارتباطاً في البناء الذي نهتم به. فإن كان قياس المحكين بثبات متساوٍ غير ممكن فإن التضليل في معاملات الصدق قد يكون أكثر ملائمة لمقارنة معاملات الصدق الملاحظة، وعلى سبيل المثال فإن تقديرات الصدق الملاحظ لعينة يساوي $\hat{\rho}_{xy} = 0.40$ ، و $\hat{\rho}_{xz} = 0.30$

ولكن تقديرات الثبات $\hat{\rho}_{xx} = 0.64$ و $\hat{\rho}_{yy} = 0.81$، و $\hat{\rho}_{zz} = 0.25$ وبتطبيق معادلة (6–10) على هذه المسألة نحصل على:

$$\rho_{t_x t_y} = \frac{0.40}{\sqrt{0.64}\sqrt{0.81}} = 0.56$$

و

$$\rho_{t_x t_z} = \frac{0.30}{\sqrt{0.64}\sqrt{0.25}} = 0.75$$

من الواضح أن استخدام المعاملات غير الملوثة قد تؤدي إلى استنتاجات مختلفة عن العلاقة بين المتغيرات عن تلك الناتجة عن استخدام معاملات صدق الدرجات الملاحظة.

الخلاصة:

يشير صدق الاختبار إلى أهمية الاستدلالات المستمدة من درجات الاختبار لهدفٍ معين ضمن مجموعة شروط وضعت مسبقاً. ويشير التصديق إلى العملية التي يتم خلالها جمع الأدلة (المؤشرات) التجريبية لتدعم استخدام درجات الاختبار للأهداف الموضوعة. وقد عرضت الأساليب الرئيسة للتصديق.

يستخدم صدق المحتوى عندما يبدو أن مستخدم الاختبار يهدف استخلاص دلالات من خلال درجات الاختبار الملاحظة في الأداء على نطاق أكبر من المهارات المشابهة لفقرات الاختبار. وعادة يتم بسؤال خبراء محكمين لفحص فقرات الاختبار والحكم على مدى تمثيل عينة الفقرات لنطاق الأداء المحدد. وفي تصميم دراسة صدق المحتوى فإنه يتم لحكم بموازنة عناصر النطاق المختلفة وتُعطى تعليمات للخبراء لمدى مطابقة الفقرات والمهام والمظاهر لتلخيص تقديرات المحكمين. وتتضمن القضايا التي قد تظهر في صدق المحتوى مدى تمثيل الأهداف للنطاق، وأهمية نطاقات معينة لمفحوصين من خلفيات عرقية أو ثقافية مختلفة، وما إذا كانت بيانات الأداء على الاختبار تناسب الحكم على صدق المحتوى.

والصدق المرتبط بمحك عبارة عن دراسة العلاقة بين درجات الاختبار ومحك أداء عملي. والصعوبات التطبيقية التي قد تظهر في مثل هذه الدراسات تتضمن تحديد المحك المناسب، وحجم العينة المناسب، وتلوث المحك، وتقييد المحك،

وغياب الثبات للمتنبئ أو المحك. وتدون نتائج الصدق المرتبط بمحك في صيغة معاملات صدق، ويمكن دعم هذه بمعامل التحديد والخطأ المعياري للتقدير وجداول التوقع، وذلك ليساعد مستخدمي الاختبارات لتقييم أهمية درجات الاختبار للأهداف المعينة.

ويكون صدق البناء مناسباً عندما يريد مستخدم الاختبار إجراء استدلالات من درجات الاختبار لنطاق سلوك لا يمكن تمثيله بمحك مفرد أو تحديده تماماً من خلال نطاق محتوى. ويتطلب صدق البناء سلسلة دراسات لاختبار فرضيات معينة حول كيفية اختلاف المفحوصين الذين يختلفون في البناء الذي نهتم به على متغيرات أخرى مرتبطة بالبناء. وقد تتضمن طرائق صدق البناء ارتباطات بين درجات الاختبار ومتغيرات محكية معينة، والفوارق بين المجموعات، والتحليل العاملي، وتحليل مصفوفة السمات – الطرائق متعددة، وعناصر تحليل التباين على أساس نظرية إمكانية التعميم. ولأن صدق البناء يمكن تطبيقه على أنواع الاختبارات جميعها ولمدى واسع من استخدامات درجات الاختبار، فإن التمييز بينه وبين الأسلوبين الآخرين للصدق قد يكون مصطنعاً نوعاً ما. والنوع الأكثر ملائمة للصدق محكومة بأنواع الاستدلالات المراد استنباطها من درجات الاختبار.

أخيراً، تم تبيان أن معاملات الصدق التي تعتمد على الدرجات الملاحظة أقل من تلك الناتجة عن الدرجات الحقيقية على مقياسين بينهما ارتباط. وتم عرض صيغة لحساب الارتباط بين الدرجات الحقيقية، وكذلك نوقشت الاعتبارات الواجب أخذها بعين الاعتبار عند استخدام معاملات الصدق.

التمارين:

1/ افترض أن مدرسة في مقاطعة كبيرة خططت لتأسيس برنامج اختبار أدنى كفاءة مطلوبة للمعلمين المبتدئين. وقد أجريت الدراسة لتقييم ملائمة أحد اختبارين منشورين على المستوى التجاري لتقرير أيهما أكثر ملائمة لتقييم الأداء على المهارات والكفاءات في هذه المقاطعة للمعلمين المبتدئين. وقام فريق من المدراء والمعلمين بمزاوجة فقرات كل الاختبارين مع قائمة المهارات المطلوبة. ووجد أن 50% من فقرات كل اختبار طابقت مهارة أو أكثر. ما البيانات الإضافية التي قد تكون مهمة في تقرير أي الاختبارين يرجح على الآخر.

2/ افترض عينة البيانات المتضمنة تقديرات ثلاثة محكمين التي تؤشر مدى مطابقة كل من الفقرات الخمس مع أهداف ثلاثة مختلفة. فهل تقديرات الفقرة الخامسة تؤشر درجة أعلى من المطابقة.

	هدف 1			هدف 2			هدف 3		
الفقرة	م1	م2	م3	م1	م2	م3	م1	م2	م3
1	1+	1+		0	0	1-	1-	1-	1-
2	1+	1+	1+	1-	1-	1-	0	1-	0
3	1+	1+	1+	1-	1-	1-	1+	1+	1+
4	1-	1-	1-	1+	1+	1+	1-	1-	1-
5	0	1-	1-	1-	0	1-	1-	1-	0

أ- ما معامل توافق الفقرة – الهدف لكل فقرة مع كل هدف؟

ب- أي الفقرات يبدو أنها تقيس الهدف الأول، الهدف الثاني، الهدف الثالث؟

ج- يبدو أن الفقرة الثالثة تطابق هدفين بشكل جيد، مع أنه ليس لها معامل توافق عالٍ مع هذه الأهداف. فسر؟ وهل تعد هذه ميزة أو عيب لهذا المعامل.

3/ افترض أن صيغتي اختبار (H وG) طورت لنطاق المحتوى نفسه، وطبقت كلا الصيغتين على المجتمع نفسه. باستخدامك الدرجات الحقيقية والدرجات الخطأ وافتراضات نظرية الدرجة الحقيقية التقليدي، اثبت أن $\sum (X_G - X_H)/N = \sigma^2_{eG} + \sigma^2_{eH}$ إذا كانت الصّيغتان متوازيتين.

4/ افترض أن مطور اختبار كُلف بإعداد صيغ عديدة لاختبار كفاءة متخصص لإجازة البورد. وكجزء من المتطلبات أجريت تجارب صدق عديدة. وأن نسبة تباين الخطأ الكلي على درجات متوسط مربعات الفروق لا تقل عن 0.80 فهل تكون هذه التجربة أكثر نجاحاً إذا كتب كل فريق من معدي الفقرات مجموعة فقرات متجانسة جداً أم إذا كتب هؤلاء مجموعة فقرات غير متجانسة؟ لماذا؟

5/ بإتباع مواصفات دراسات صدق عدة. اقرأ كل وصف مما يأتي، وحدد أي المشكلات يمكن أن تؤثر في النتائج:

أ- أعد معلم رياضيات في المرحلة الثانوية اختبار لتحديد الطلبة المبدعين، وطبق الاختبار على (100) طالب في المدرسة، وعلى أساس درجات الاختبار العالية

يسمح للطلبة الانخراط في مساقات الرياضيات في كلية جامعية محلية. وفي نهاية المساق حسب مطوّر الاختبار الارتباط بين درجات الاختبار مع درجاتهم في مساق الكلية. وقد وجد المعلم أن الارتباط له دلالة إحصائية.

ب- طوّر عالم نفس صناعي اختبار لتعيين متقدمين إلى وظائف مهنية في شركة معينة وطلب من المشرفين على المهن تحديد مواقع العاملين الحاليين على تدريج مؤلف من فئات: أعلى من الوسط ووسط ودون الوسط، وبعدها اختبر العاملين الذين كانوا من فئة أعلى من الوسط وفئة أدنى من الوسط، وقارن بين متوسطاتهم، ووجد عالم النفس أن الفرق ذا الدلالة الإحصائية يحابي مجموعة أعلى من الوسط.

ج- اقترح عالم نفس اجتماعي اختبار في النضج الاجتماعي للأطفال. وطبق لاختبار على (50) طفل في مرحلة ما قبل المدرسة، وعلق الوالدان والمعلمون على درجات هؤلاء الأطفال. وقد طلب من المعلمين ملاحظة الأطفال بالاهتمام للأشهر الثلاثة القادمة والاحتفاظ بسجلات لسلوكات كل طفل التي قد تفيد في تقييم مستوى النضج الاجتماعي، ورتب عالم النفس الأطفال بالاعتماد على مستوى النضج الاجتماعي، ورتبهم أيضاً بالاعتماد على درجات الاختبار. وحسب معامل ارتباط الرتب لسبيرمان وكان عالياً.

د- تم إجراء صدق لاختبار ذكاء جمعي وذلك من خلال ارتباطه باختبار ستانفورد - بينيه. وبعدها طور عالم نفس تربوي صيغة مختصرة للاختبار الجمعي الذي يتضمن فقرات لفظية فقط، وحسب الارتباط بين درجات الصيغة المطوّلة والصيغة المختصرة، وجد أنه عالٍ وموجب. واقترح بعدها استخدام الصيغة المختصرة للاختبار الجمعي ليحل محل اختبار ستانفورد - بينيه والذي يتطلب وقتاً أطول ليطبق على كل مفحوص.

6/ مع الأخذ بعين الاعتبار نوع الاستنتاج الذي يرغب به مطور الاختبار من درجات الاختبار، ولكل من المواقف الآتية، بين أيّ نوع من دراسات الصدق أكثر ملائمة (صدق المحتوى أم الصدق المرتبط بمحك تنبؤي، أم الصدق المرتبط بمحك تلازمي أم صدق البناء)؟

أ- عالم نفس اجتماعي طور قائمة بهدف التعريف بالعاملين بالوظائف المرتبطة بالصحة. واستخدمت العديد من السلوكات لتحديد "الاحتراق النفسي" المتضمن الغياب العالي، الفشل في التوافق مع الأمن المؤسسي، والعلاقات المشدودة مع الزملاء والمشرفين.

ب- طور عالم نفس تطوري مختص بالمراهقة استبانة لتطبق على طلبة المدرسة المتوسطة للتنبؤ بمن سيفشل في المدرسة الثانوية.

ج- تم تطوير تدريج لتقدير الأداء الصفي للمعلمين في السنة الأولى في التعليم، وطلب من المعلمين الخبراء فحص الفقرات على التدريج المقدّر لرؤية ما إذا كانت تلائم الأداء الفعّال.

د- طور اختبار استعداد أكاديمي غير لفظي للأطفال ذوي إعاقات السمع في مرحلة ما قبل المدرسة. واستخدم الاختبار كمقياس للانتقاء لتحديد الأطفال الذين لديهم صعوبات تعليمية في القراءة بالفصول العادية.

7/ فيما يأتي مصفوفة السمات - الطرائق المتعددة أُعدت من بيانات دراسة ماركس ووابن (Marx & Winne, 1978). افحص مصفوفة الارتباطات هذه واجب عن الأسئلة التي تليها، ودوّن مؤلفي هذه الدراسة الارتباط المصحح من أثر التلوث (وقد عدَّ / مؤلفو الدراسة الاختبارات الفرعية على أنها السمات وأدوات مفهوم الذات المختلفة على أنها الطرائق المستخدمة في قياس هذه السمات).

الطريقة أ				الطريقة أ			
1	2	3	4	1	2	3	4

أ- قائمة جوردن لمفهوم الذات

1. المظهر الفيزيائي	0.70			
2. الاجتماعي	0.64	0.64		
3- الملائمة الشخصية	0.97	0.75	0.73	
4- الملائمة الأكاديمية	0.60	0.88	0.92	0.66

ب- قائمة بيرز - هاريس

1- المظهر الفيزيائي	0.78	0.66	0.63	0.50	0.80			
2- الشعبية	0.59	0.59	0.62	0.42	0.87	0.79		
3- القلق	0.53	0.57	0.55	0.52	0.72	0.70	0.73	
4- الدراسات الاجتماعية	0.59	0.81	0.68	0.80	0.90	0.86	0.75	0.80

اقتبست عن:

W. Marx and P.H, winne Construct Interpretation of three Self - concept inventories. American Educational Research Journal, 15 , 99-109, Table 2 ، Cop-

yright1978 , American Educational Research Assn, Washington, D.G.

أ- ما تقديرات الثبات للاختبارات الفرعية الأربعة لكل أداة.

ب- ما قيم معاملات الصدق التجميعي.

ج- أي من السمات الأربع يظهر أن لها أكبر معامل صدق تجميعي.

د- أي من معاملات السمات المختلفة - الطريقة الواحدة تطرح تساؤلات حول صدق البناء لسمة المظهر الفيزيائي.

هـ- أي من معاملات السمات المختلفة - الطرق المختلفة تطرح تساؤلات حول صدق البناء لسمة الدراسات المدرسية/ الأكاديمية؟

8/ افترض أن الباحث اختار أن يدون معاملات الصدق غير الملوثة لبيانات المصفوفة المبينة في السؤال السابع. ما القيم الممكنة لمعاملات الصدق التجميعي.

9/ باستخدام المعادلة (10-6) في حساب معامل الصدق غير الملوث. بين اشتقاق مختصر للعلاقة:

$$\rho_{xy} \leq \sqrt{\rho_{xx'}} \sqrt{\rho_{yy'}}$$

10/ لاحظ طالب دراسات عليا في علم النفس التحليلي في دليل اختبار شخصية أن معامل ثبات درجات الاختبار = 0.76 ، ولكن معامل الصدق لهذا الاختبار واختبار آخر يقيس سمة مشابهة = 0.78 ، وتوقع أن هنالك خطأ لأنه سمع "أن الاختبار أكثر صدقاً من ثباته". فهل هذا صحيح بالضرورة؟ لماذا نعم ولماذا لا؟.

11/ افترض أنه تم تطوير اختبار جديد في الاستعداد لطلبة الدراسات العليا. ويريد مطور الاختبار أن يكون معامل ارتباط هذا الاختبار مع متوسط درجات الفصل الدراسي الأول بمعامل صدق لا يقل عن 0.60 وحصل مطور الاختبار من دراسة استطلاعية على تقدير ابتدائي للثبات لمجموعة الفقرات الابتدائية. اشتق صيغة لحساب الحد الأدنى لـ (K) للمعامل الذي يجب أن يزداد بنسبته طول الاختبار للحصول على معامل الصدق المطلوب. وهل أن زيادة طول الاختبار بنسبة المعامل (K) يؤكد تحقيق معامل الصدق المرغوب؟ فسر أهمية مثل هذه الصيغة؟

الفصل الحادي عشر

11

الطرائق الاحصائية للتنبؤ والتصنيف

الفصل الحادي عشر

الطرائق الاحصائية للتنبؤ والتصنيف

عرضنا في الفصل السابق مفاهيم صدق الاختبار ومعامل الصدق البسيط في إثبات قوة العلاقة بين درجات الاختبار والمحك. وفي معظم المشكلات التي تستخدم التنبؤ بالأداء الفردي في المواقف التحليلية، الصناعية، والتربوية يتوافر لدى الباحث عدد من المتنبآت المناسبة. وأحد أهداف هذا الفصل هو وصف الطرائق الإحصائية واسعة الانتشار لتصديق استخدام اثنين أو أكثر من المتغيرات في التبنؤ بالأداء على محك مفرد ومصحح على تدريج متصل. وفي هذا السياق سنصف طرائق الارتباط الجزئي والانحدار المتعدد. وهدف ثاني هو وصف الطرائق الإحصائية شائعة الاستخدام في تصنيف المفحوصين عندما تكون قيم المحك محددة على تصنيف متقطع (من فئتين أو أكثر). ويعد تحليل دوال التمييز تقنية مستخدمة في مثل هذه المواقف. وقد طرحت الطرائق الإحصائية هنا بدرجة كافية من التعقيد وذلك لتسويغ الحصول على كتاب معالجته شاملة للموضوع.

والعرض في هذا الفصل مقيدٌ بتفسير مختصر للطريقة، وتم عرض مثال لدراسة الصدق التي تستخدم الطريقة وعرض النتائج مثالياً مع التحليل والتفسير المرافق لها. والقصد من هذا العرض تزويد طلبة نظرية القياس بالمعرفة الأساسية الضرورية لتمييز متى وكيف تطبق هذه الطرائق في جهودهم الخاصة لتصديق الاختبارات ولفهم تقارير دراسات الصدق الموجودة في أدبيات القياس التي تستخدم هذه الطرائق، ويمكن حذف المادة المطروحة في هذا الفصل دون فقدان الاستمرارية مع الموضوعات التالية المغطاة في هذا الكتاب.

الارتباط الجزئي:

هنالك مواقف في صدق الاختبار يتوافر لدى مستخدم الاختبار واحدٌ أو أكثر من المتنبآت المفيدة، ولكن نطرح الآن السؤال: ما قوة العلاقة بين المحك ومتنبئ إضافي لمجموعة فرعية من المفحوصين المتجانسين بالنسبة للمتنبئ المحدد مسبقاً؟ على سبيل المثال افترض لمشكلة[1] القبول المدرسي في القانون والتي يقبل بوساطته الطلبة على أساس متوسط درجاته في

(1) دون بور في (Powers, 1982) معاملات الصدق LSAT واختبار القبول في كلية الحقوق، في هذه المسألة.

الدراسة الأولية التي تستخدم كمتنبئ لمتوسط الدرجات في السنة الأولى لدراسة القانون. افترض أن الارتباط بين هذا المتنبئ والمحك $\hat{\rho}_{yx}=0.46$ ، ومتنبئ أخر ممكن هو درجة المقبولين في اختبار القبول لمدرسة القانون (LSAT) والذي ارتباطه مع متوسط درجات السنة الأولى في دراسة القانون $\hat{\rho}_{yx2} = 0.48$ وقد يكون المشرف على القبول مهتماً في معرفة قوة العلاقة بين LSAT ومتوسط GPA في مدرسة القانون وذلك للطلبة المتشابهين في GPA في الدراسة الأولية (أي المتجانسين بالنسبة لـ X_1: ومعامل الارتباط الذي يفيد في هذا الموقف هو معامل الارتباط الجزئي الذي يرمز له بـ $\hat{\rho}_{yx2x1}$، ويشير الرمز السفلي إلى أننا مهتمون بالعلاقة بين Y و X_2 لمجموعة متجانسة من التلاميذ بالنسبة لـ X_1. وصيغة حساب معامل الارتباط الجزئي هذا:

$$\frac{\hat{\rho}_{yx1}-\hat{\rho}_{yx2}\;\hat{\rho}_{x1x2}}{\sqrt{1-\hat{\rho}^2_{yx1}}\;\sqrt{1-\hat{\rho}^2_{x1x2}}} \qquad (1-11).............$$

وكما يتبين من هذه الصيغة فإن قيمة الارتباط الجزئي لا يعتمد على الارتباط بين Y و X_2 فقط بل وعلى ارتباط Y و X_1 والارتباط بين المتنبئين X_1 و X_2، فعلى سبيل المثال في مسألة مدرسة القانون إذا كان الارتباط بين GPA الأولية ودرجات LSAT حيث $\rho_{x1x2} = 0.33$، فإن الارتباط الجزئي المحسوب يكون:

$$\hat{\rho}_{yx2x1} = \frac{0.48 - (0.46)\,(0.33)}{\sqrt{1-(0.46)^2}\;\sqrt{1-(0.33)^2}} = 0.39$$

ومن جهة أخرى إن كان الارتباط بين المتنبئ الأول والثاني أكبر (وليكن $\rho_{x1x2} = 0.60$) فإن الارتباط الجزئي بين y و X_2 مع ضبط X_1 يكون:

$$\hat{\rho}_{yx2x1} = \frac{0.48 - (0.46)\,(0.60)}{\sqrt{1-0.46^2}\;\sqrt{1-0.60^2}} = 0.29$$

ومعامل الارتباط هذا أقل من معامل الارتباط الأول وذلك لأن الارتباط بين المتنبئين في هذه الحالة عالي وموجب. وأخيراً يجب ملاحظة أن الارتباط الجزئي يمكن أن يتجاوز الترابط الأصلي إذا كان الارتباط بين X_1 و X_2: $\hat{\rho}_{x_1x_2}=-0.25$، فإن قيمة $\hat{\rho}_{yx2\text{-}x1}$ يكون:

$$\hat{\rho}_{x2yx1} = \frac{(0.25)\,(0.46) - 0.48}{\sqrt{1 - 0.25^{2}}\ \sqrt{1 - 0.46^{2}}} = +0.6$$

وفي هذه الحالات جميعها فإن القيمة الموجبة لمعامل الارتباط الجزئي بين LSATوالمتوسط في مدرسة القانون GPA يمكن تفسيره بشكل عام ليعني أنه وللمفحوصين المتشابهين في المتوسط بالدراسة الأولية، يجب أن تُعطى الأولوية للذين حصلوا على درجات أعلى على LSAT. علاوة على ذلك يوضح المثالين الأول والثاني أنه عندما يكون ارتباط كلا المتنبئين إيجابي مع المحك فإن مستخدميه أو فائدة المتنبئ الثاني تنخفض كلما اقترب ارتباط المتنبئين من 1.

ويمكن توسيع الارتباطات الجزئية لتجزئة أي عدد من المتغيرات، وبالطبع عند توافر عدد كافي من المفحوصين. (واثبت تجريبياً أن حجم عينات دراسات الارتباط هو العدد الأكبر من الآتية: 100 فرد أو 10 إضعاف عدد متغيرات الدراسة). ومع أن حساب معامل الارتباط الجزئي ممكن يدوياً، إلا أن وصف العملية الحسابية معقد وحذفت الصيغ الحسابية هنا، ويمكن إيجادها في بيدهازر (**Pedhazur, 1982**)، وعادة تجرى هذه الحسابات بوساطة الحاسوب.

ومن المهم ملاحظة أنه كلما تم تثبيت متغيرات إضافية فإن الارتباط قد يتغير في الإشارة أو القيمة، ومن المهم أيضاً تمييز أثر خطأ القياس في الترابط الجزئي، والقيمة المطلقة للارتباط من الدرجة صفر (الارتباط دون أي تجزئة) بين المتغيرات المشوشة بخطأ القياس أقل من القيمة المطلقة للارتباط الذي نحصل عليه للبيانات الثابتة تماماً. والارتباط الجزئي لمتغيرات مشوشة بخطأ القياس قد يكون أكبر أو أصغر من أو مختلف في الإشارة عن الارتباط الذي تم الحصول عليه لبيانات ثابتة تماماً، وصيغة تصحيح الارتباط الجزئي لمحك وفئتين اثنتين الناتج عن التشويش يكتب على النحو الآتي:

$$\rho_{yx2.x1} = \frac{\hat{\rho}_{yx2} - \hat{\rho}_{11} \quad \rho_{yx1}\,.\,\hat{\rho}_{x1x2} -}{\sqrt{\hat{\rho}^{2}_{yx1} - \hat{\rho}_{11}.\,\hat{\rho}_{yy}}\quad \sqrt{\hat{\rho}^{2}_{x1y2} - \hat{\rho}_{11}.\,\hat{\rho}_{22}}} \qquad (2\text{-}11)\ldots\ldots\ldots\ldots\ldots$$

حيث ترمز ρ_{11}، ρ_{22}، ρ_{yy} إلى تقديرات الثبات. افترض أن تقديرات الثبات لمتوسط طلبة الدراسة الأولية، ودرجات LSAT، ومتوسط مدرسة القانون هي 0.67,0.87,0.64 على التوالي، والارتباط الجزئي الحقيقي المقدر بين LSAT ومتوسط السنة الأولى GPA في مدرسة القانون (وذلك بعد التجزئة لـ GPA في الدراسة الأولية):

$$0.51 = \frac{(0.33)\,(0.46) - (0.48)\,(0.64)}{\sqrt{0.33 - (0.87)\,(0.64)}\ \sqrt[2]{0.46 - (0.67)\,(0.64)}} = \hat{\rho}_{yx2\text{-}x1}$$

ومن المهم تذكر أن الارتباط الجزئي الحقيقي المحسوب قد يكون مفيداً لأهداف نظرية (في صدق البناء)، إلا أنه في المواقف العملية يجب أن يعمل مستخدم الاختبار مع مقاييس تنبؤ ومحك ثباتها ليس تاماً (أي أقل من الثبات التام). ويهتم المثال السابق باستخدام الارتباط الجزئي في سياق صدق التبنؤات. وعندما تكون الارتباطات الجزئية بينGPA في السنة الأولى ودرجات LSAT دالة إحصائياً يمكننا القول أن LSAT له صدق تنبؤي عند إضافته إلى متوسط GPA في الدراسة الأولية كمتنبئ لمتوسط السنة الأولى في مدرسة القانون. ويمكن استخدام الارتباط الجزئي في سياق صدق البناء، على سبيل المثال افترض لمفهوم الذات، يميز بعض المنظرين بين مفهوم الذات الأكاديمي العام ومفهوم الذات بالنسبة لموضوعات أكاديمية معينة، وتتوافر اختبارات لقياس كل منهما، فالنظرية تقترح أن التحصيل في موضوع أكاديمية معين يرتبط إيجابياً مع كل من مفهوم الذات الأكاديمي العام ومفهوم الذات بالمجال الأكاديمي الخاص. إضافة على ذلك فإن النظرية تقترح أن مفهوم الذات الخاص بمجال أكاديمي معين يجب أن يرتبط إيجابياً بالتحصيل في المجال نفسه حتى للمجموعات المتجانسة في مفهوم الذات الأكاديمي العام. ويجب أن يتذكر القارئ بأن التحقق من صدق البناء يستخدم تحديد ما إذا كانت درجات الاختبار تتطابق مع التنبؤات النظرية. وعندما يحدث التطابق فإنه يعد مؤشراً إيجابياً لصدق بناء الاختبارات المستخدمة.

وقد جمع شافيلسون وبولس **(Shavelson & Bolus, 1982)** بيانات اختبار تحصيلي في العلوم (Y) ومفهوم الذات الأكاديمي في العلوم (X_1) ومفهوم الذات الأكاديمي العام (X_2) مع متغيرات أخرى. وكانت الترابطات بين أزواج المتغيرات الثلاث على النحو الآتي:

$$0.73 = \rho_{x1x2} \quad ، \quad 0.41 = \rho_{yx2} \quad ، \quad 0.49 = \rho_{yx1}$$

ويكون الارتباط بين التحصيل في العلوم ومفهوم الذات الأكاديمي في العلوم (بتجزئة مفهوم الذات الأكاديمي) كما يأتي:

$$\frac{\hat{\rho}_{yx2}\,.\,\hat{\rho}_{x1x2} - \hat{\rho}_{yx1}}{\sqrt{\hat{\rho}_{x1x2} - 1}\ \sqrt{\hat{\rho}^2_{yx2} - 1}} = \rho_{yx11\ x2}$$

$$= \frac{(0.73)\,(0.41) - 0.49}{\sqrt{1-(0.73)^2}\,\sqrt{1-(0.42)^2}} = 0.31$$

ويدعم هذا الترابط الادعاء بأن صدق البناء لمقياس مفهوم الذات في العلوم يتفق مع التبنؤات النظرية، ويجب ملاحظة أن الارتباطات المدونة من قبل شافيلسون وبولس يجب تعديلها لإزالة أثر الخطأ من أثر التشويش (التوهين).

الانحدار المتعدد:

عرض الفصل الثاني الانحدار المتعدد كمنهجية تفيد في تطوير معادلة تنبؤ عند توافر درجات متنبئ ودرجات محك واحد لمجموعة المفحوصين نفسها. وعند توافر اثنين أو أكثر من المتنبآت يكون الانحدار المتعدد مفيداً في تطوير معادلة واحدة للتنبؤ بالأداء المحكي من مجموعة المتنبآت، تذكر الصيغة العامة لمعادلة التنبؤ لمتبنئ واحد:

$$Y' = C + b_{y.x}\,X \qquad \text{.............(3-11)}$$

وترمز X إلى درجة المفحوص على المتنبئ، و Y' إلى درجة المفحوص المتنبأ بها على المحك، و $b_{y.x}\ \mu_{by.x} - \mu_y =$ ، وتذكر أيضاً أنه حالما تستخدم درجات عينة لحساب C و $b_{y.x}$ فإنه يمكن استخدام معادلة الانحدار للحصول على قيمة متنبأ بها لـ Y (أي مفحوص) عند توافر درجته على X، مع العلم بأن المفحوص والعينة المستخدمة في حساب معادلة الانحدار اختيرتا من المجتمع نفسه، وإن كانت القيمة المطلقة للارتباط بين x و y عالية لدرجة كافية فإن القيم المتنبأ بها لـ y تكون دقيقة وغير متحيزة (وبالرجوع إلى معادلة 2 –21، والمثال الحسابي الذي يليها يفيد في مراجعة هذه المفاهيم).

مثال بمتنبئين اثنين:

مع أن هنالك مواقف يتوافر فيها لمستخدم الاختبار متنبئ واحد، والأفضل إيجاد متنبئين أو أكثر لاستخدامها في محاولة التنبؤ بأداء المفحوصين أو وصفه، بالمقابل فإن استخدام متغير واحد على أنه العامل الوحيد في اتخاذ القرار يعد عملاً غير حكيم (طائش). للتوضيح خذ الموقف الذي يهدف إلى تحديد مجموعة متنبآت تفيد في تحديد الأطفال المتخلفين في التطور المعرفي عند عمر 3 سنوات، فمن المرغوب به هنا اعتماد المتنبآت على معلومات متوافرة عند الولادة. وفي هذا الموقف هنالك متنبئين ممكنين هما الوزن عند الولادة والدرجة على تدريج لعوامل المخاض والولادة تحدد من خلال تاريخ الحالة أثناء الحمل والولادة. وقبل

الاستخدام الفعلي لهذه المتنبآت فمن الضروري تطوير معادلة تنبؤية لجمع درجات الوزن عند الولادة وعوامل المخاض والولادة. ومن الضروري أيضاً تحديد مدى دقة المتغيرين في التنبؤ على المحك. ومعادلة التنبؤ بالأداء على المحك من متنبأين هي:

$$X_2 b_{yx2.x1} + X\, b_{yx1.x2} + c = y\backslash \qquad (4\text{-}10)..........$$

وفي مثالنا تكن`y هي القيمة المتنبأ بها على مقياس ستانفورد – بينيه، وترمز x_1 إلى الوزن عند الولادة، و x_2 الدرجة على متغير عوامل المخاض والولادة والحدين $b_{yx1.x2}$ و$b_{yx2.x1}$ يسميان معاملات الانحدار و C هي القاطع، والقيم العددية لمعاملات الانحدار والقاطع نحصل عليها بحل المعادلات باستخدام المتوسطات والانحرافات المعيارية للمتغيرات والارتباطات عبر أزواج المتغيرات. ويبين جدول (11–1)[2] هذه الإحصائيات لمثال التطور المعرفي والصيغ لمعاملات الانحدار والقيم العددية لهذا المثال على النحو الآتي:

$$\frac{(\hat{\rho}_{x1x2}\,\hat{\rho}_{yx2} - \hat{\rho}_{yx1})\,\hat{\sigma}_{y}}{(\hat{\rho}^{2}_{x1x2} - 1)\,\hat{\sigma}_{x1}} = b_{yx1.x2} \qquad (5\text{-}11)..........$$

$$0.006 = \frac{(0.111\text{-})\,(0.149)\ \text{-}(0.301)\ 19.251}{[^{2}(0.111\ \text{-})\ \text{-}1\ ]\ 887.641} =$$

$$\frac{(\hat{\rho}_{yx1}\,\hat{\rho}_{y1x2} - \hat{\rho}_{yx2})\,\hat{\sigma}_{y}}{(\hat{\rho}^{2}_{xx2} - 1)\,\hat{\sigma}_{x2}} = b_{yx2.x1} \qquad (6\text{-}11)..........$$

$$0.247 = \frac{(0.111\text{-})\,(0.301)\ \text{-}(0.149)\ 19.251}{[^{2}(0.111\ \text{-})\ \text{-}1\ ]\ 9.121} =$$

---

(2) الشكر لجون هولسترم (June Holstrum) لسماحه لنا استخدام بيانات المثال مع ملاحظة ان هذه البيانات جمعت كجزء من دراسة هدفها أكثر شمولا مما ذكرنا.

جدول (11-1): المتوسطات، الانحرافات المعيارية، ومعاملات الارتباط لمقياس ستانفورد بينيه، ووزن الولادة، وعوامل المخاض والولادة.

	Y	X_1	X_2
Y ستانفورد بينيه	1.00		
X_1 وزن الولادة	0.301	1.00	
X_2 عوامل الولادة والمخاض	0.149-	0.111-	1.000
المتوسط	95.326	2724.900	6.445
الانحرافات المعيارية	19.251	887.641	9.121

وزن الولادة مقاس بالغرامات.

والصيغة والقيمة العددية للقاطع هي:

$$B_{yx2\,x1}\,\mu_{x2} - b_{yx1x2}\,-\mu_{y} = C \qquad (7\text{-}11)\ldots\ldots\ldots$$

$$(6.445)\ (0.247\text{-}) - (-2724.9)\ (0.006) - 95.326 =$$

$$80.568 =$$

وبالتعويض لقيم القاطع ومعاملات الانحدار المحسوبة بمعادلة التنبؤ العامة (معادلة 11-4)، نحصل على معادلة تنبؤ خاصة بهذه الدراسة:

$$X_2\ 0.247 - X1\ .006\ +\ 80.568 = y\backslash$$

لطفل وزنه عند الولادة 2000 غم ولعامل المخاض والولادة 20، فإن درجة IQ المتنبأ بها على مقياس ستانفورد – بينيه تكون:

$$(20)\ .247 - (2000)\ 0.006\ + 80.568\ = y\backslash$$

$$87.6 =$$

وكما في الانحدار الخطي، فإن قيم القاطع C ومعاملات الانحدار (b'S) التي حصلنا عليها بهذه الطريقة ستقلل من الفجوة بين درجات y الفعلية على المحك والدرجات المتنبأ بها (y\).

تذكر أنه في الانحدار البسيط يرتبط معاملي الانحدار والارتباط بعلاقة بسيطة على النحو الآتي:

$$\hat{P}_{yx} \quad \frac{\hat{\sigma}_y}{\hat{\sigma}_x} = b_{yx} \qquad (8-11)..........$$

وبطريقة مماثلة، فإن معاملات الانحدار والارتباط الجزئي تربط بينهما علاقة بسيطة، والعلاقة بين $b_{yx2.x1}$ و $P_{yx2.x1}$ على النحو الآتي:

$$\hat{P}_{yx2x1} \frac{\hat{\sigma}_{y.x1}}{\hat{\sigma}_{x2.x1}} = b_{yx2.x1} \qquad (9-11 أ)..........$$

وهنا يرمز $\hat{\sigma}_{y.x1}$ إلى الخطأ المعياري للتنبؤ بـ y منX_1، و $\hat{\sigma}_{x2.x1}$ إلى الخطأ المعياري للتنبؤ بـ X_Z منX_1 . ومع أن هنالك علاقة بسيطة بين الارتباط الجزئي ومعامل الانحدار، إلا أنهم يعبرون عن خصائص مختلفة للبيانات. فمعامل الارتباط الجزئي $\hat{P}_{yx1.x2}$ يعبر عن قوة العلاقة الخطية بين Yو X_2 لمجموعة مفحوصين متجانسين على X_1 بينما يعبر معامل الانحدار $b_{yx2.x1}$ عن كمية التغير عندما تتغير X_2 نقطة واحدة وبقاء X_1 ثابتة، لذلك فهي تعبر عن حساسية $Y^{\backslash}$ للتغير. أما العلاقة بين $b_{yx1.x2}$ و $\hat{\rho}_{yx1.x2}$ فهي على النحو الآتي:

$$\hat{\rho}_{yx1x2} \frac{\hat{\sigma}_{y.x2}}{\hat{\sigma}_{x1.x2}} = b_{yx1.x2} \qquad (9-11 ب)..........$$

وتستخدم علاقات مشابهة للعلاقات في المعادلات (11–9 أ) و (11– 9 ب) عندما يكون هنالك أكثر من متنبئين. ويتم حساب معاملات الانحدار من درجات IQ الخام والوزن عند الولادة وعوامل المخاض والولادة. ومع أن الدرجات الخام شائعة الاستخدام في تحليل الانحدار، إلا أن الباحثين يستخدمون الدرجات الزائية (درجات Z) بدلاً من الدرجات الخام، وعند استخدام الدرجات الزائية لتحل محل الدرجات الخام فإن معادلة الانحدار تكتب على النحو الآتي:

$$\beta_{yx2.x1} Z_{x2} + \beta_{yx1.x2} Z_{x1} = Z^{\backslash}y \qquad (10-11).............$$

وهنا ترمز β'S إلى معاملات انحدار محسوبة باستخدام الدرجات المعيارية. لاحظ أن معادلة (11–10) لا تتضمن رمز للقاطع، وذلك لأن القاطع يساوي صفراً دائماً عند استخدام الدرجات الزائية في تحليل الانحدار. ويربط معاملات الانحدار المعيارية والمعاملات غير المعيارية علاقة بسيطة، وعلى النحو الآتي:

$$\hat{b}_{yxixj} \quad \frac{\hat{\sigma}_{xi}}{\hat{\sigma}_y} = \beta_{yx1.xj} \qquad (11-11).............$$

ويبين هذا التعبير أن ضرب المعامل غير المعياري لمتنبئ ما بنسبة الانحراف المعياري للمتنبئ إلى الانحراف المعياري للمحك. وللتوضيح سنحسب الانحراف المعياري للوزن عند الولادة:

$$\hat{b}_{yx1x2} \quad \frac{\hat{\sigma}_{x1}}{\hat{\sigma}_{y1}} = \beta_{yx1.x2}$$

$$0.006 \times \frac{887.641}{19.251} =$$

$$0.276 =$$

وتطبق العلاقة في المعادلة 11–11 عندما يكون هنالك أكثر من متنبئين اثنين.

وسيدهش القارئ فيما لو كانت المعاملات المعيارية لها مزايا على المعاملات غير المعيارية. والميزة التي يدعيها البعض هي أن القيم العددية لمعاملات معيارية اثنين يمكن مقارنتها مباشرة في حين قيم المعاملات غير المعيارية لا يمكن مقارنتها مباشرة. لذا فإن المعاملات المعيارية يمكن استخدامها لاختيار المتنبئ الأكثر فائدة. لماذا حدث هذا الادعاء؟ سبب هذا هو لمقارنة وحدات القياس، فمعامل الانحدار لـ X_j هي الكمية التي تتغير بها القيمة المُتنبأة لـ Y عندما تتغير X_J وحدة واحدة مع بقاء X'S الأخرى ثابتة، وهذا لأن معامل الانحدار غير المعياري للوزن عند الولادة = 0.006، فلطفلين الفرق بين وزنيهما 1غم ولهما درجات عوامل المخاض والولادة نفسها سيختلفان عن بعضهما البعض بـ 0.006 نقطة على درجاتIQ المتنبأ بها، وبسبب هذا التفسير لمعاملات الانحدار فإن أي مقارنة عددية لمعاملين تتطلب تطبيق وحدات قياس المتنبآت في مقارنة المعامل غير المعياري للوزن عند الولادة (0.006) مع المعامل غير المعياري لعوامل المخاض والولادة (0.247)، ويجب أن نفترض هنا أن وحدة قياس واحدة على تدريج وزن الولادة مكافئاً لوحدة واحدة على تدريج عوامل المخاض والولادة. ومن الواضح أنه لا يمكن الدفاع عن هذا الافتراض، وللتخلص من هذه المشكلة يحسب بعض الباحثين معاملات الانحدار المعيارية ويقارن فيما بينها. والافتراض هنا أن درجات Z لها وحدة القياس نفسها للمتغيرات جميعها. على سبيل المثال في التنبؤ لـ I.Q من وزن الولادة وعوامل المخاض والولادة، يتطلب مقارنة معاملات الانحدار المعيارية افتراض أن درجات Z لكلا العاملين لها وحدة القياس نفسها، إذ أن وحدة واحدة على تدريج Z مكافئة لانحراف معياري واحد على تدريج الدرجات الخام. ويتضمن هذا الافتراض أن انحراف معياري واحد

(887.64 غم) على تدريج وزن الولادة مكافئ لانحرافاً معيارياً واحداً (9.12 نقطة) على تدريج عوامل المخاض والولادة، يحتاج هذا النوع من الافتراضات الموازنة الدقيقة قبل مقارنة الأوزان المعيارية. ويبدو أن هذا الافتراض غير واضح المعالم في المثال الحالي في أحسن الأحوال. ومع ذلك ففي المواقف التي تكون المتنبآت هي مقاييس لبناءات تربوية أو نفسية فإن هذه الافتراضات يمكن الدفاع عنها بدرجة أكبر.

تقويم دقة التنبؤ:

يعد تقويم دقة نتائج معادلة التنبؤ أهمية أولية في دراسة الصدق، ويمكن استخدام إحصائيات عدة لوصف دقة التنبؤ والتدوين العلمي لها وفقا لما يمليه الضمير، وتتضمن عادة معاملات لكلا النوعين الآتيين:

1. قياس ارتباطي لدقة التنبؤ.

2. قياس لخطا التنبؤ المحتمل في تقدير درجات أداء الأفراد على المحك.

وهنالك إحصائيات عديدة لكل فئة سيتم وصفها في الأجزاء الآتية.

القياس الارتباطي لدقة التنبؤ: لتطوير قياس ارتباطي لدقة التنبؤ من المهم التمييز بين معادلة تنبؤ العينة:

$$b_{yx2.}X_2 + b_{yx1.x2}X_1 + C = y^{\backslash}$$

ومعادلة التنبؤ للمجتمع:

$$\beta_{yx2.x1}X_2 + \beta_{yx1.x2}X_1 + C = y^{\backslash}$$

والمعادلة الأخيرة هي التي نحصل عليها فيما لو حصلنا على درجات مجتمع المفحوصين بأكمله فكل مفحوص له ثلاثة قيم مختلفة نهتم بها هي:

Y: درجة المحك الفعلية.

$Y^{\backslash}$: درجة المحك المتنبأ بها باستخدام عينة الانحدار.

$Y^{\backslash\backslash}$: درجة المحك المتنبأ بها باستخدام معادلة انحدار المجتمع.

لتطوير مقياسنا لدقة التنبؤ، خذ المواقف الافتراضية الآتية: افترض أنه يمكن استخدام معادلة تنبؤ العينة لحساب Y لكل فرد في المجتمع ونحصل فيها على $Y^{\backslash}$. وبعدها نحسب مربع معامل الارتباط بين Y و $Y^{\backslash}$ للمجتمع، ويعبر مربع معامل الارتباط هذا عن دقة $Y^{\backslash}$

كمتنبئ لـY الذي يسمى مربع معامل ارتباط الصدق المتقاطع Squared cross- validated correlation coefficient" ونرمز له بـρ^2_{cv}: مقياس الارتباط لدقة التنبؤ. ويتضمن مصطلح معامل الصدق المتقاطع معادلة تنبؤ العينة التي تم فحصها لمجموعة غير تلك التي اشتقت منها،وهو أن ρ^2_{cv} تخبرنا عن مدى أهمية معادلة تنبؤ العينة عند تطبيقها على مفحوصين آخرين في المجتمع. ومع أن ρ^2_{cv} ذات أهمية أولية في دراسة الصدق فمن المهم فهم أن مقدار هذا المعامل يرتبط مباشرة بمقدار معاملين إضافيين اثنين: الأول هو مربع معامل الارتباط المتعدد (ρ^2)والذي يحدد على أنه مربع الارتباط بين y و\Y. والمعامل ρ^2 يقيس قوة الارتباط بين المحك ومجموعة من المتنبآت، والثاني هو مربع معامل الارتباط $\rho_{yy'}$ بين قيم y و\y وقد بين روزبوم (Roozeboom,1981) أنَّ ρ^2_{cv}: $\rho = \rho^2, \rho^2_{yy'}$ لذا فإن قيمة ρ^2_{cv} تعتمد على عاملين اثنين هما:

أولاً: تعتمد على ρ^2 والذي يقيس دقة التنبؤ بـ y إذا ما استخدمنا معادلة المجتمع، كذلك تقيس مدى قوة ارتباط المتنبآت مع المحك.

ثانياً: ρ^2_{cv} وتعتمد على $\rho^2_{y1.y2}$، أي مقياس للتشابه بين معادلتي تنبؤ العينة والمجتمع، ويعتمد $\rho_{y1,y}$ بشكل أولي على النسبة بين حجم العينة إلى عدد المتنبآت، فإن كانت هذه النسبة صغيرة فإن $\rho^2_{y'y}$ تكون صغيرة أيضاً، وبالتالي دقة تنبؤ ضعيفة كما قيست بـρ^2_{cv} الناتجة إما عن ضعف العلاقة بين المحك والمتنبآت أو البيانات غير الكافية لحساب الصدق مع المتنبآت العديدة، فمن المرغوب فيه حساب كل من ρ^2، ρ^2_{cv} بمتنبئين، ويمكن حساب ρ^2 بالمعادلة:

$$R^2 = \frac{\hat{\rho}_{yx2}\ b_{yx2.x1}\ .\ \hat{\sigma}_{x2} + \hat{\rho}_{yx1}\ b_{yx1.x2}\ .\ \hat{\sigma}_{x1}}{\hat{\sigma}_y}$$

وفي مثالنا تكون قيمة R^2 كما يأتي:

$$R^2 = \frac{(0.111)\ (0.247\text{-})\ 9.121+ (0.301)\ (0.006)\ 887{,}641}{19.251}$$

$$= 0.10$$

وحيث يأن هذا التقدير لـ ρ^2 منخفض جداً، وأن ρ^2_{cv} تكون أقل منρ^2، فمن غير الضروري حسابρ^2_{cv} في هذا المثال. ونحن نعرف أنه قد يكون صغيراً، وهذا لأن ارتباط الوزن وعوامل

المخاض والولادة ضعيف مع مقياس سيتانفورد – بينيه عند عمر 3 سنوات. لذا وبغض النظر عن عدد المفحوصين في مثالنا، فإن معادلة التنبؤ لا تزودنا بتنبؤات دقيقة. ومع ذلك سنستخدم المثال نفسه في توضيح حساب ρ^2_{cv}.

وقد طور براوني (Browne, 1975) طريقة لحساب ρ^2_{cv}، وتتطلب كمتيان (R_c^4 له R_c^2)على أنها عناصر R^2_{cv} أولاً:

$$R_c^2 = \frac{(R^2 - 1)\, k - R^2}{1 - k - N} \qquad (14\text{-}11)\ldots\ldots\ldots\ldots$$

حيث ترمز k إلى عدد المتنبآت وتساوي 2 في مثالنا، و R_c^2 إلى تقدير ρ^2 والتي تعد أقل تحيزاً من R^2 تقريباً، ويشار إليها عادة على أنها R^2 المعدلة.

إضافة إلى ذلك يجب أن نحسب قيمة :Rc^4

$$Rc^4 = \frac{{}^2(R_c^2 - 1)\, K2\ - (R_c^2)^2}{(1+K-N)(1-N)} \qquad (15\text{-}11)\ldots\ldots\ldots\ldots$$

واستخدام الرمز R_c^4 من قبل براوني ولا يشير إلى أنه مربع R_c^2 ، ويحسب مربع معامل ارتباط الصدق المتقاطع من خلال:

$$Rcv^2 = \frac{R_c^2 + R_c^4\,(3 - K - N)}{K + R_c^2\,(2 - K_2 - N)} \qquad (16\text{-}11)\ldots\ldots\ldots\ldots$$

وفي مثالنا، حساب هذه الكميات يؤدي إلى:

$$R_c^2 = \frac{(0.1 - 1)\, 2 - 0.1}{1 - 2 - 101} = 0.081.$$

$$R_c^4 = \frac{{}^2(0.081 - 1)\ (2)2 - {}^2\,(0.081)}{(1+2-101)\,(1 - 101)} = 0.006$$

$$\text{و}\quad R^2_{cv} = \frac{(0.081 + (.006)\ (3\text{-}2\text{-}101)}{2+(0.081)\ (2\text{-}\ (2)\ (2)\ \text{-}\ 101)} = 0.067$$

وكما يتبين من المسافات فإن R^2_{cv} قليلة ومشابهة في قيمتها لـ R^2، وتشير هذه ثانية إلى غياب دقة التنبؤ، وذلك لأن وزن الولادة ودرجة الاختبار الحركي النفسي (السيكوموتور) ليست متنبآت قوية لمقياس ستانفورد – بينيه، وليس لعدم كفاية البيانات لحساب معادلة التبنؤ، وعلى القارئ ملاحظة أن $R_c{}^2$ يمكن أن تكون سالبة، وإن كانت صفراً يجب تعويض $R_c{}^2$ بصفر في المعادلات (11 – 15) و (11 – 16). كذلك فإن $R_c{}^4$ يمكن أن تكون سالبة، وإن كانت صفراً فإنها تعوض بصفر في المعادلة (11 – 16).

مقارنة بنتائج المثال السابق لنأخذ توضيحاً آخر لاستخدام R^2، R^2_c و R^4_{cv} اعتماداً على نتائج دراسة هارتلي وهارتلي (**Harthley & Harthley, 1976**)، وقد جمعا بيانات (57) طالباً انخرطوا في مساق مناهج البحث في علم النفس، وتألفت البيانات من (22) متنبئا واختبار تحصيلي لهذا المساق. وقد دونا تقدير مربع الارتباط المتعدد لـ R^2 (0.86) (ويبدو أنه يشير إلى علاقة قوية بين المحك والمتنبئ). ومع ذلك فإن استخدام المعادلة (11–14) لحساب R^2 المعدلة نحصل على $R^2_c = 0.77$ ، لذلك فإننا نرى أن قيمة R^2 الأصلية مضللة نوعاً ما لقياس قوة العلاقة. وبدون مقدمات وجدنا أن R^2_c دونت على أنها تقدير لمربع ارتباط الصدق التقاطع(R^2_{cv})، ولكن يمكننا ومن خلال مثال هارتلي وهارتلي رؤية أن هذا يكون مضللاً. وعند تطبيق المعادلتين (11–15) و (11– 16) على بيانات هارتلي وهارتلي نجد أن $R^2_{cv} = 0.64$ في هذا المثال. ومن الواضح أن انطباعنا عن دقة الصدق التنبؤي تتغير دراماتيكيا بالانتقال من R^2 إلى R^2_{cv}، لذا فإن هذا المثال يوضح أهمية تقديرات مربع معامل الارتباط المتعدد للصدق التقاطع. بالإضافة إلى أن R^2_{cv} أقل بشكل ملحوظ من R^2_c، وهذه تقترح أن جمع بيانات مفحوصين إضافيين يفيد في هذه الدراسة في زيادة الدقة التنبؤية لمعادلة انحدار العينة.

وهنالك طرائق أخرى لحساب R^2_{cv}، أحداها تستخدم دراسة صدق تقاطعي، والأسلوب الشائع هنا هو استخدام نصف عينة المفحوصين لحساب معادلة التنبؤ، وتطبيق المعادلة فيما بعد على النصف الآخر من البيانات للحصول على قيم Y' لهؤلاء المفحوصين. ويحسب بعدها معامل الارتباط البسيطة بين y و y' للحصول على تقدير لـ ρ^2_{cv} وتكمن مشكلة هذه الطريقة في التضحية بنصف البيانات في حساب معادلة التنبؤ.

مقاييس خطأ التنبؤ:

يمكن تدوين دقة التنبؤ بصورة مباشرة أكثر نوعاً ما من R^2_{cv} باستخدام الخطأ المعياري للتقدير والخطأ المعياري لـ $y^\backslash$. ومع أن الأولى أقل مرغوبية كمقياس لدقة التنبؤ نوعاً إلا أنه يمكن أخذها بالاعتبار أولاً لأنها أكثر شيوعاً.

وقد تم تقديم الصيغة الرياضية لحساب الخطأ المعياري للتقدير لمتنبئ واحد مع مثال رياضي في الفصل الثاني، وثانية مع تفسير الاستخدام في دراسات الصدق في الفصل التاسع. ويحسب الخطأ المعياري للتقدير لعدة متنبآت بالمعادلة:

$$\hat{\sigma}_{\underline{x}.y} = \sqrt{(R^2 - 1)\ \ \sigma^2_y\ \frac{1 - N}{K - N}} \qquad (17\text{-}11)\ldots\ldots\ldots\ldots$$

ويشير الخطأ تحت x إلى وجود أكثر من متنبئ، و R^2 تمثل مربع الارتباط المتعدد المحسوب. وفي مثال دراسة صدق الأطفال الرضع، يكون تقدير قيمة الخطأ المعياري كما يأتي:

$$\hat{\sigma}_{x.y} = \sqrt{(0.1 - 1)\ (370\ .601)\ \frac{100}{99}} = 18.355$$

ومع أن σ_{xy} واسعة الاستخدام في تقدير خطأ التنبؤ إلا أنها قيمة مُغالى فيها لدرجة دقة التنبؤ، وذلك لأنها لا تعكس حقيقة أنه ولأي مفحوص $Y^\backslash$ (القيمة المتنبأ بها من بيانات العينة) تعد فقط تقدير لـ $\dot{y}$ (القيمة المتنبأ بها من بيانات المجتمع).

ويمكن كتابة خطأ التنبؤ $E = Y\text{-}\dot{Y}$ على النحو الآتي:

$$E = (\dot{y} - y) + (y^\backslash - \dot{y}^\backslash) \qquad (18\text{-}11)\ldots\ldots\ldots\ldots$$

حيث يعكس الفرق $(Y\text{-}\dot{Y})$ الفجوة بين Y الحقيقية والتنبؤ بـ y من بيانات المجتمع، وتعكس $(y^\backslash\text{-}y)$ الفجوة بين قيمة y المتنبأ بها من بيانات المجتمع وقيمة Y المتنبأ بها من بيانات العينة ويتضمن الخطأ المعياري لـ $y^\backslash$ كلا من مصدري الخطأ للمتنبئين الاثنين، الذي هو مربع الخطأ المعياري ويحسب من

$$SE^2_{y^\backslash} = \frac{N\text{-}1}{N\text{-}2}\ (R^2\text{-}1)\ \hat{\sigma}^2_y \left[+ 1 + \frac{1}{N} + \frac{1}{(\rho_{x1x2} - 1)\ (N\text{-}1)}\ \ \frac{1}{(\rho_{x1xc1Z2}\ 2\text{-}\ Z_2^2 + Z_1^2)} \right]$$

(19 -11).............

جدول (11-2): الأخطاء المعيارية لـ $Y^\backslash$ لتجمعات Z1 و Z2.

Zz					
2	1	0	-1[a]	-2[a]	Z_1
19.1	18.9	18.8	18.9	19.2	2-
18.9	18.6	18.5	18.6	18.9	1-
18.8	18.5	18.4	18.5	18.8	صفر
18.9	18.6	18.5	18.6	18.9	1
19.2	18.9	18.8	18.9	19.1	2

$0.111- = \hat{\rho}_{x1x2} =$ ، 101 =N

a: قيم Z_2 = - 1 أو -2 لا تظهر في البيانات الفعلية التي اعتمد عليها هذا الجدول.

وتمثل $\hat{\rho}_{x1x2}$ هنا الارتباط بين X_1و X_2 و $Z_1 = \frac{\hat{M}_1 - X_1}{\sigma_{X1}}$ و $Z_2 = \frac{\hat{M}_2 - X_1}{\sigma_{X2}}$ و N عدد المفحوصين في العينة. ويبين جدول (11-2) قيمة هذا الإحصائي لتجمعات متنوعة من الدرجات في الوزن عند الولادة والاختبار السيكومتري للأطفال الرضع. وبتفحص جدول11-12 يتبين لنا أن التنبؤ يكون أكثر دقة للمفحوصين الذين درجاتهم قريبة من المتوسط على كلا المتنبئين وأقل دقة للمفحوصين الذين تبعد درجاتهم عن متوسطات كلا المتنبئين. وفي هذا المثال لا يتغير الخطأ المعياري لـ $y^\backslash$ دراماتيكياً بتغير Z_1 و Z_2، وذلك لانخفاض قيمة ρ_{x1x2} والكبر النسبي لحجم العينة، وهذا لا يظهر في حالات أخرى. ويبين الجدول (11-3) قيم الخطأ المعياري لـ y التي قد تظهر مع $\rho_{x1x2} = -0.707$ و N = 30 وفي هذه الحالة يكون أثرZ_{x1}وZ_{x2} أكثر دراماتيكياً .

ويمكن استخدام الخطأ المعياري لـ$y^\backslash$ في تكوين فترات تنبؤ للمفحوصين، فعندما يكون حجم العينة أكبر من 30، فإن الفاحص يمكن أن يبني فترة $y^\backslash + S.Ey^\backslash$ وبتأكيد نسبته68% بان قيمة y الحقيقية تقع ضمن هذه الفترة.ومن المدهش ملاحظة أن التقدير المعياري لـ$y^\backslash$ يتأثر بشكل أساسي بقيم $\hat{\rho}_{x1x2}$ و N وفي الوقت نفسه عدم تأثر الخطأ المعياري للتقدير المحسوب بالمعادلة (11-7) بخصائص العينة هذه، وبالتالي فانه يعدّ غير دقيق.

جدول (3-11): الأخطاء المعيارية لـ $Y^{\backslash}$ لتجمعات Z_1, Z_2

Z_2					
2	1	0	-1	-2	Z_1
18.9	18.8	19.2	19.8	20.8	2-
18.8	18.5	18.6	19.1	19.8	1-
19.2	18.6	18.4	18.6	19.2	صفر
19.8	19.1	18.6	18.5	18.8	1
20.8	19.8	19.2	18.8	18.9	2

$0.707 = \hat{\rho}_{x_1x_2}$ ،$30 = N$

التنبؤ باستخدام أكثر من متنبئين اثنين:

يمكن كتابة معادلة تنبؤ العينة عندما يستخدم أكثر من متنبئين اثنين على النحو الآتي:

$$X_k\, x_{k-1} \,.....\, b_{yxk}\, X_1 + ... + X_1 x_1 \,......\, b_{yx1x2} + C = Y^1 \quad (20-11)....$$

ويرمز C إلى القاطع و b's لمعاملات الانحدار لمتغيرات عددها K، والحرفين على يسار النقطة أسفل معامل الانحدار تحددّ المحك والمتنبئ الذين سيطبق عليهما المعامل. والحروف إلى يمين النقطة تشير إلى المتنبآت الأخرى في المعادلة. ومثال محسوس لمعادلة التنبؤ لأكثر من متغيرين، افترض أنه بالإضافة إلى الوزن عند الولادة (X_1)والدرجات على عوامل المخاض والولادة (X_2) في المثال السابق، لدينا بيانات عن عمر الأم عند ولادة الطفل (X_3) والحالة الاجتماعية الاقتصادية للأم(X_4)، وبالتالي فإن معادلة انحدار العينة تكون:

$$by_{x4.\,x1x2x3}\, X4 + by_{x3.\,x1x2x4}\, X3 + by_{x2.\,x1x3x4}\, X2 + by_{x1.\,x2x3x4}\, X1 + C = Y1$$

وبوجود أكثر من متنبئين، فإن حساب معاملات الانحدار يدوياً يكون معقداً، وعادة يجري باستخدام برنامج محوسب، والشيء نفسه ينطبق على:

$\hat{\rho}_{y.y}$ و $S.E\ Y^{\backslash}$ و R^2

اختيار المجموعة الفرعية للمتنبئ:

عندما يتوافر اثنان أو أكثر من المتنبآت، فمن المهم الأخذ بالاعتبار ما إذا كانت مجموعة المتنبآت الفرعية ستعطي دقة تنبؤية قريبة من تلك لمجموعة المتنبآت جميعها. بالمقابل فإن مجموعة فرعية منها يمكن أن تؤدي إلى دقة تنبؤ أكثر من تلك الناتجة عن المجموعة الكلية (Hocking, 1976). والسبب في محاولة إيجاد مثل هذه المجموعة الفرعية لإلغاء جمع بيانات على متنبآت تساهم بقليل أو لا تساهم بدقة التنبؤ. فعلى سبيل المثال قد يكون هنالك ارتباطاً عالياً لمجموعتين مع محك الأداء، وفي هذه الحالة فلا داعي لتضييع الوقت والجهد والكلفة في تطبيق كليهما.

افترض وفي مثالنا السابق أن الباحث اختار مجموعة فرعية من المتنبآت الفعّالة للقياس ولها دقة تنبؤية معقولة. وهنالك العديد من التقنيات الممكنة لاختيار المتنبئ ولكن تقنية واحدة تعرف بـ: انحدارات المجموعات الفرعية الممكنة جميعها، وهذه سيتم عرضها هنا. والمهتم بتفاصيل أكثر لهذه التقنية وتقنيات اختيار المتغير الأخرى يمكنه الرجوع إلى هوكنج (Hock-ing, 1976) والذي قدم مراجعة واسعة للطرائق والقضايا المتعلقة في اختيار المتغيرات. وكما يشير الاسم يستخدم انحدار المجموعات الفرعية الممكنة جميعها في حساب دقة التنبؤ لكل تجمع ممكن للمتنبآت. وتستخدم هذه النتائج فيما بعد في الإرشاد لاختيار مجموعة ملائمة من المتنبآت. ويعرض جدول (11–4) نتائج تحليل انحدار المجموعات المحتملة جميعها لمتغيرات الوزن عند الولادة وعوامل المخاض والولادة وعمر الأم وحالة الأم الاجتماعية الاقتصادية كمتنبآت لمقياس ستانفورد – بينيه.

ويدون الجدول قيم R^2 لكل تجمع محتمل لمتنبئ واحد، أو اثنين أو ثلاثة أو أربعة.

وتشير هذه النتائج إلى أن R^2_c لنموذج المتغيرات الأربعة أنها أكبر قليلاً من R^2 للعديد من نماذج المتغيرين الاثنين، لذا فإن النموذج بمتغيرين اثنين من المتنبآت الأربعة المحتملة يبدو مناسباً. ولأن الفروق قليلة بين إحصائيات R^2 لأفضل ثلاثة نماذج المتغيرين الاثنين، وأن SES مستخدم فيها جميعاً، لذا يقترح نتيجة ذلك أن الجمع بين أي من المتغيرات الثلاثة المتبقية مع SES تزودنا باختيار مناسب عبر المتنبآت الأربعة. كذلك تم تدوين قيم معاملات Rc في الجدول نفسه لإثبات أنه وبزيادة عدد المتغيرات (K) فإن R^2_c تصل إلى قمة قصوى ثم تنخفض، كذلك فإن R^2_{cv} سيصل أعلى قيمة عند قيمة K نفسها.

وكنتاج: قانون معقول لاختيار المتغيرات هو اختيار عدد من المتنبآت التي ينتج عنها أكبر قيمة لـ R^2_c، ويشير هذا القانون ثانية إلى اختيار متنبئَين اثنين.

جدل (11-4): نتائج تحليل الانحدار للمجموعات الفرعية المحتملة جمعيها.

الرقم في المعادلة	R2	Rc2	المتغير في المعادلة
1	0.03		IF
1	0.08		MAGE
1	0.09		BW
1	0.24	0.23	SES
2	0.09		BW,IF
2	0.11		MAG,IF
2	0.15		BW,MAGE
2	0.25		MAGE, SES
2	0.27		IF, SES
2	0.32	0.32	BW,SES
3	0.15		BW, MAGE
3	0.27		MAGE, IF,SES
3	0.32		BW,IF,SES
3	0.32	0.30	BW,MAGE, SES
4	0.33	0.30	BW,MAGE,IF,SES

تحليل التمييز:

بمجموعتين:

نواجه أحياناً في دراسات الصدق تطوير طريقة لاستخدام بيانات الاختبار في تصنيف المفحوصين إلى مجموعتين أو أكثر. فعلى سبيل المثال افترض أن مستشفى معين صنف مرضاه إلى فئتين شيزوفرانيا وكآبة، اعتماداً على مقابلات تحليلية. وتتطلب عملية المقابلة عدداً كبيراً من المحللين والكثير من وقتهم، فقد يكون من الأهمية بمكان التحقق من صدق بطارية اختبارات فرعية تقيس اضطرابات التفكير "Thinking disturbance" والتخلف "withdrawal distursbance"والانسحاب العدواني "hostile suspiciousness“ وقلق الاحباط ”anxious depression"وقدم كل من كلينباوكوبر كوبر,anxinbaum & Kupper) (1978 بيانات افتراضية للحالة الموصوفة أعلاه.

جدول (11-5): بيانات توضيحية للتحليل التمييزي:

	المتغير				
مجموعة التحليل	اضطرابات التفكير X1	إعاقة X2	الانسحاب العدواني X3	قلق الاحباط X4	دالة التمييز Y
الكآبة	5.0	6.4	6.6	10.2	-3.41
	2.9	5.3	2.5	9.2	-2.62
	2.7	5.0	2.5	10.3	-3.01
	2.5	4.7	3.5	8.6	-3.01
	1.9	4.3	4.5	11.1	-0.84
الشيزوفرانيا	7.8	6.9	3.8	6.0	-12.70
	7.6	5.6	4.2	7.3	-14.34
	7.4	5.1	5.4	4.9	-20.12
	7.2	6.0	6.0	3.6	-8.42
	3.5	5.6	3.0	11.8	-0.33

مقتبس عن د. ج. كلينباوم و ل. ل. كوبر (1978). تطبيق تحليل الانحدار وطرائق متعددة المتغيرات (North Scituate: Duxbury Press)

وفي الجدول (11-5) فإن مجموعات التحليل النفسي هي المجموعات التي صنف بوساطتها المرضى من قبل المحللين النفسيين والدرجات على x1 وحتى x4 تمثل البيانات الملاحظة على المتغيرات الأربعة. وتم حساب الدرجات على yمن الدرجات على x1وحتى x4 ، وكانت الطريقة المتبعة على النحو الآتي: يبدأ تحليل التمييز من درجات المفحوصين في العينة الذين تصنيفهم الصحيح معروف. واستخدمت هذه الدرجات لحساب أوزان المتغيرات المجمعة في متغير واحد استخدم على الفور في تصنيف المفحوصين الذين تصنيفهم الصحيح غير معروف. وفي المثال الحالي، فهذا يعني أن الدرجات على X_1 وحتى X_4 ستستخدم لتحديد أوزان التمييز d,c,b,a وذلك للمجموع الخطي الموزون:

$$d_{x4} + c_{x3} + b_{x2} + a_{x1} = y \qquad \text{(21-11)} \ldots\ldots\ldots\ldots$$

ويعرف الجمع الخطي هذا باسم دالة التمييز، وحالما تتحدد قيم d,c,b,a فإنها تستخدم في حساب قيمة y للمرضى الذين لم يتم مقابلتهم من قبل المحلل النفسي. وعلى القارئ أن يلحظ وجه الشبه بين الاستخدام المقصود لتحليل التمييز، واستخدامنا في البداية لتحليل الانحدار المتعدد. ففي الأخير كان الهدف استخدام العينة لتحديد معادلة يمكن استخدامها للتنبؤ بدرجات المحك للمفحوصين مستقبلاً، بينما في الأولى كان الهدف تطوير معادلة يمكن استخدامها في تصنيف المفحوصين مستقبلاً. وبما أن y تستخدم في تصنيف المفحوصين، فإنه يجب تكوين مجموعتين مختلفتين لأقصى قدر ممكن على y. وقياس طبيعي للفرق بين المجموعتين هو:

$$w = \frac{\hat{\mu}_{yL} - \hat{\mu}_{yi}}{\hat{\sigma}y} \qquad (22\text{-}11)...........$$

وتمثل الكمية y الفرق في المتوسط المعياري بين المجموعتين (الكآبة والشيزوفرانيا)، والمحك المعتمد في اختيار أوزان d,c,b,a وكذلك w2 أن تكون أكبر ما يمكن.

جدول (11-6): المعاملات لدوال التحليل النفسي والتمييز

تصنيف مقابلة التحليل النفسي	دالة تصنيف التمييز الكآبة	الشيزوفرانيا
الكآبة	5	0
الشيزوفرانيا	1	4

والطريقة الرياضية المستخدمة في تحديد الأوزان ليست من أهداف هذا الكتاب، وللمهتم يمكن أن يجدها في كليناوم وكوبر (1980).

وفي مثالنا كانت الأوزان كما يأتي:

a = -3.22 ، b= 1.61 . c= 1.15، d = -0.51 لذا فإنه وللمريض الأول (الكآبة). فإن y تكون :

$$(10.22)\ (0.51\text{-}) + (6.6)1.15 + (6.4)\ 1.61 + (5.0)\ (3.22\text{-}) = y$$
$$3.41 =$$

والدرجات على معادلة دالة التمييز للمرضى الآخرين مبينة في جدول (11–5). وواضح أن مرضى الكآبة تكون درجاتهم أعلى من مرضى الشيزوفرانيا فمتوسط درجة دالة التمييز لمرضى الكآبة –2.0 ، ولمرضى الشيزوفرانيا –11.2. وهذا يقترح أن هنالك قاعدة منطقية لتصنيف المرضى تستخدم متوسط المتوسطين ومن ثم تصنيف أي مريض من خلال درجة دالة التمييز، وتصنيف أي مريض دالة التمييز له أعلى من الوسط على أنه مريض كآبة، ومتوسط المتوسطين هو:

$$-6.6 = 2\ /\ [(2.0\text{-}) + 11.2\text{-}]$$

وعلى هذا فإن أي مريض درجته أعلى من –6.6. يصنف على انه مريض كآبة والآخرين يصنفوا على أنهم مرضى شيزوفرانيا[1].

وحالما يتم تطوير القانون فمن المهم تقويم دقته، وطريقة بسيطة مصاحبة لهذا هو استخدام القانون في تصنيف المرضى الذين طور القانون من خلالهم. وحيث أن المرضى الذين درجاتهم أعلى من (–6.6) صنفوا على أنهم مرضى كآبة والآخرين على أنهم مرضى شيزوفرانيا، وهذه الطريقة صنفت الجميع إلى مرضى كآبة عدا الأخير فكان شيزوفرانياً. ويمكن عرض هذه النتيجة في جدول مثل جدول (11–6)، واعتماداً على هذا الجدول فان احتمالية التصنيف المحسوب لمرضى الكآبة = صفر وللشيزوفرانيا =0.20 .

والمشكلة في تقويم قانون التصنيف مع البيانات التي اعتمدت عليها هو احتمال أن تكون الدقة مغالى فيها. وهنالك أسلوبان للتغلب على هذه المشكلة. الأول هو تطبيق القانون على مجموعة مرضى لم تستخدم في تطوير القانون وإحدى الإجراءات المستخدمة هي تطوير القانون من خلال نصف البيانات المتوافرة وتقييمه من خلال النصف الآخر. وهذه الطريقة مشابهة لطريقة الصدق المتقاطع الموصوفة في تقويم دقة التنبؤ لمعادلات الانحدار المتعدد، وتشترك معها في أنه يتم إهمال نصف البيانات. والثاني مشابه لطريقة براوني في تقدير الصدق المتقاطع، وقد وصف كل من كلينباوم وكوبر(1978) وموريسون (Morrison, 1976) مثل هذه الطرائق.

---

(3) يستخدم متوسط المتوسط على افتراض انه وفي المجتمع الذي سيتم تطبيق القانون عليه، تتوافر حالات كثيرة لكلا المرضين وبالتساوي. وقد وصف كلينباوم وكوبر (1978) قوانين مناسبة عندما يكون هذا الافتراض غير صحيح.

وفي المثال السابق كان الهدف تطوير طريقة تصنيف، مع اهمام قليل بالتفسير الحقيقي لأوزان التمييز.مع ذلك افترض أن البطارية طورت لتقويم فاعلية العلاج السايكايتري (التحليل النفسي)، والنظرية التي يقترحها تطوير الاختبار هو أن البطارية يجب أن تميز بين مرضى الكآبة ومرضى الشيزوفرانيا. بالإضافة أنه يمكن تمييز هاتين المجموعتين بدقة في اضطراب التفكير وقلق الإحباط مع ممارسة اضطراب التفكير بدرجة اكبر وقلق الإحباط بدرجة أقل لمرضى الشيزوفرانيا وفي هذا الموقف يوجد نوع من الاهتمام في تفسير y، إذ إن y هي المتغير الذي يميز بأقصى قدر بين المجموعتين، وتفسير y مرتبط بصدق بناء البطارية.

وعندما يكون هدف التحليل تفسير الفرق بين المجموعتين فإن هنالك العديد من الإحصائيات المناسبة. وقد قدم بري وماكسويل (Bray & Maxwell,1982) مراجعة كاملة لهذه الإحصائيات. وسيتم هنا فقط عرض للإحصائيات التي يعتقد أنها الأكثر أهمية في توضيح المعنى الأساسي لدالة التمييز.والطريقة الأكثر شيوعاً لاستخدام دالة التمييز في تفسير الفروق بين المجموعات هو فحص أوزان التمييز، والاستدلال بوساطة هذا الأسلوب يمكن رؤيته من خلال إعادة النظر لدالة التمييز في المثال الحالي، معادلة دالة التمييز هي:

$$X_4d + X3C + X2b + X1a \ =Y \qquad (23\text{-}11)\text{.............}$$

ومن الواضح أن المتغير ذا أكبر وزن تمييزي له تأثير أكبر فيY، وذلك أن التغير بقيمة معينة على هذا المتغير تؤدي إلى اكبر تغيير في Y. وللتوضيح تذكر أن التعويض لـd,c,b,a يؤدي على:

$$X_4\,(0.51^{-}) + X_3 1.15 + X_2\ 1.61 + X_1\ \ 3.22\text{-} = y$$

وبالتالي فإن التغير نقطة واحدة على X_1 تؤدي إلى أكبر تغير في y من التغير نقطة واحدة على X_2 أو X_3 أو X_4، لذا فإن X_1 تفسر على أنها لها أكبر تأثير في y، و X_4 لها أقل تأثير. والاستنتاج سيكون هو نفسه بغض النظر عن قيمة التغير الحاصل في الـ X'_s

والمشكلة في هذا الأسلوب أنه يفترض أن التغير نقطة واحدة على الـ X'_s من بعض الوجوه مكافيء للتغير نقطة واحدة على أي من الـ X'_s الأخرى. بمعنى آخر أنها تفترض تساوي وحدة القياس للمتغيرات جميعها، ومن الواضح أن هذا الافتراض خاطئ في معظم التطبيقات.

ولمحاولة الإحاطة بهذه المشكلة فان عدداً من الفاحصين حسبوا الأوزان التي قد تنتج فيما لو أجري تقنين أو معايرة لكل X، وهذه الأوزان تعرف بأوزان التمييز المعيارية، واعتبرت عند العديد أنها قابلة للمقارنة من متغير لآخر.

وفي مثالنا الأوزان المعيارية هي -4.89، 1.17 ، 2.56، و -1.18 وعلى أساس هذه الأوزان فإن اضطراب التفكير بقي المساهم الأكبر في دالة التمييز، مع أن مساهمته بالنسبة للإعاقة والشك العدواني كان أكبر في حالة الأوزان المعيارية عنه في الأوزان الأصلية. ومن جهة أخرى فإن مقدار مساهمته بالنسبة لقلق الإحباط حكم على انه أقل.

وكبديل للأوزان المعيارية، يمكن فحص الارتباط بين المتغيرات ودالة التمييز على أنها الأساس في تفسير دالة التمييز.

وباستخدام X_1 لتوضيح حساب الارتباطات فإن صيغة الارتباط بين y و X_1 تكون.

$$\rho_{14\,d} + \rho_{13c} + \rho_{12\,b} + a = \rho_{y}\quad (24\text{-}11)$$

حيث تمثل ρ'S على الجهة اليسرى الارتباط المحسوب بمعالجة جميع البيانات كما لو أنها مجموعة واحدة. لذا فإنه لحساب الارتباط بين دالة التمييز والمتغير فإن الارتباط بين هذا المتغير والمتغيرات الأخرى جميعها تضرب بالوزن التمييزي للمتغير الآخر، وتجمع هذه النتائج فيما بعد ويضاف إليها وزن المتغير قيد الدراسة. وفي مثالنا كان الارتباط بين y و X_1 حتى X_4 على النحو الآتي: -0.89 ، – 0.34 ، -0.35 ، 0.78 على التوالي . لذا فإن دالة التمييز لمرضى الشيزوفرانيا تميل لأن تكون درجاتها منخفضة، وتظهر ارتباطاً سالباً مع اضطراب التفكير وايجابي مع قلق الإحباط

وفي الاختيار بين أوزان التمييز المعيارية وارتباطات دالة التمييز- أوزان التمييز فمن المهم التمييز بين هذين الإحصائيين والذي يجيب كل منهما عن أسئلة مختلفة (Bray & Maxwell 1982), وتلخص أوزان التمييز التأثير المفرد للمتغير على دالة التميز، في حين أن الارتباطات تعبر عن التباين المشترك بين المتغير ودالة التمييز، ووافق كل من بري وماكسويل على أن الارتباطات تعد مرشداً أفضل للتفسير الجوهري لدالة التمييز.

مع أكثر من مجموعتين:

يمكن توسيع استخدام تحليل التمييز لموقف تستخدم فيه مجموعات عدة. فعندما يكون لدينا مجموعتان فقط فإن دالة تمييز واحدة تعزى للفروق جميعها بين المتغيرات ومع ذلك سنهتم بدراسة الفروق عبر عدة مجموعات، وسيتبين لنا أن الفروق بين المجموعات لا تكون مقدمة بشكل مناسب في دالة التمييز الأولى، فبعض المجموعات المختلفة بالفعل بالمتغيرات الأصلية قد تظهر متشابهة على دالة التمييز الأولى. وفي مثل هذا الموقف نحتاج لأكثر من دالة تمييز واحدة لتفسير الفروق بين المجموعات، وفي هذه الحالة فان دالة التمييز الثانية عبارة

عن الجمع الموزون للمتغيرات الأصلية، والتي من خلال التجمعات الموزونة الممكنة غير المرتبطة بدالة التمييز الأولى، وهذه تعظم الفروق عبر المجموعات. وستكون دالة التمييز الثالثة عبارة عن المركب الموزون وغير المرتبط مع أي من الدالتين الأولى والثانية، وهذه تزودنا بأقصى فصل بين المجموعات. وأكبر عدد ممكن من دوال التمييز يساوي عدد المتغيرات p أو أقل بواحد من عدد المجموعات والتي هي أقل. وقد قدم كل من كلينباوم وكوبر(1978) وموريسون(1976) طريقة حساب دالة التمييز عند توافر أكثر من مجموعتين.

وقدم بري وماكسويل (1982) مثالاً لاستخدام تحليل التمييز المتعدد لتفسير الفروق عبر أربع مجموعات من الأطفال صنفت على أنها ذات:

(1)صعوبات قرائية شديدة.

(2) صعوبات قرائية متوسطة.

(3) قارئين متوسطين.

(4) قارئين متميزين. ولهؤلاء الأطفال تم الحصول على درجات على المتغيرات المحكية الآتية:

اختبار القراءة اللفظية المصور(PPVT)، اختبار الذكاء اللفظي العام، اختبار الطلاقة اللفظية (VF)، اختبار المشابهات على مقياس وكسلر لمرحلة ما قبل المدرسة والمرحلة الابتدائية (SIM)، اختبارين في القدرة القرائية، اختبار التكامل الحركي البصري (VMI)، واختبار المعرفة –التمييز(R)، وكلاً من الإختبار غير اللفظي العام واختبار الشكل (EF) والتي تقيس القدرات غير اللفظية العالية. أظهر تحليل التمييز دالتي تمييز لها دلالة إحصائية، ومن خلال الأوزان المعيارية للتمييز على أنها المعاملات، فإن دالتي التمييز كانتا على النحو الآتي:

$$Y = 0.15\ (PPVT) + 0.10\ (RD) + 0.39\ (EF) + 0.33\ \text{و}\ (VF) + 0.29\ (VMI) + 0.44\ (SIM)$$

$$Y = 0.95\ (PPVT) + (-0.51)\ (RD) + (-0.58)\ (EF) + 0.37\ (VF) + 0.12\ (VMI) + (-0.2)\ (SIM)$$

وتشير الأوزان المعيارية للتمييز على أن المتغيرات (SIM,VF,EF) هي الأكثر أهمية في الفصل بين المجموعات في الدالة الأولى. بينما في الدالة الثانية كان الوزن الأكبر لـ PPVTومساهمة عالية لا وزن كل من RD و EFو VF .

ويبين جدول (11–7) ارتباطات دالة التمييز، فدالة التمييز الأولى لها ارتباط موجب مع

المتغيرات جميعها، والعلاقات الأقوى كانت لـ EFو SIM، والعلاقة بين دالة التمييز الأولى والمتغيرات المتبقية متساوية في قوتها تقريباً. ودالة التمييز الثانية كان ارتباطها موجب مع كل من PPVT و VFوسالب معRD وEF، وبشكل أساسي لا ارتباط مع VMIوSIM. ويظهر متغير التمييز الثاني المقابلة بين قدرات التمييز اللفظية والبصرية.

جدول (11-7): معاملات دوال التمييز لمثال بري - ماكسويل.

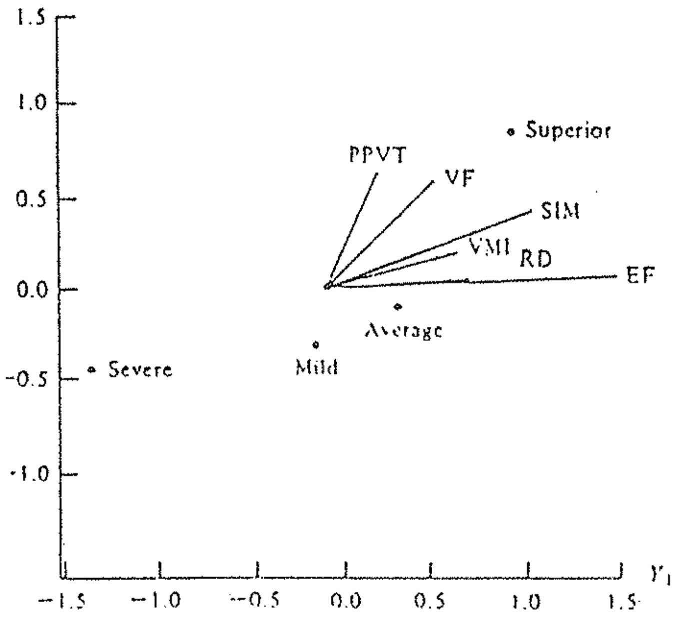
المتغير	دالة التمييز	
	الأولى	الثانية
PPVT	0.46	0.25
RD	0.52	-0.37
EF	0.73	-0.36
VF	0.53	0.33
VMI	0.50	-0.10
SIM	0.64	0.04

مأخوذ عن بري وماكسويل. تحليل وتفسير تحليل التباين المتعدد. Reviewal Educe. Research, 52.

وعندما يكون هناك أكثر من دالة تمييز معنوية، فإن متوسط دوال المجموعات يمكن تمثيله بيانياً كدليل في تفسير دوال التمييز وطبيعة الفروق بين المجموعات. ويمثل الشكل (11-1) رسم بياني لمتوسط المجموعة على دوال التمييز لمجموعات القراءة الأربع، والنقطة الموضوعة مع اسم المجموعة تمثل متوسطات درجات تلك المجموعة على دالتي التمييز. وهذه النقطة على التمثيل البياني يطلق عليها اسم مجموعة المركز "Group Centroid"وهذه النقطة والخطوط المنبثقة عنها هي مركز العينة الكلية. ومن الواضح أن دالة التمييز الأولى تفصل كل مجموعة عن غيرها من المجموعات، وتميز دالة التمييز الثانية القارئين المتميزين عن المجموعات الثلاث الأخرى، ومن الممكن تمثيل القياسات على المتغير الأصلي من خلال متجهات، لذا فإنه يمكن تكوين تفسير لمدى الاختلاف بين المجموعات على هذه المتغيرات. مثل هذه المتجهات تشير إلى المجموعات ذات أعلى مستويات متوسط على المتغيرات الأصلية، وبعيداً عن المجموعات التي لها أدنى مستويات متوسط. ويشير طول كل متجه إلى فعاليته التمييزية عبر المجموعات (انظر Overall & Klett لبيان طريقة الرسم) ويمكننا رؤية ومن خلال موقع المتجهات أن PPVT

يميز بشكل أساسي بين القارئين المتميزين عن المجموعات الأخرى، وEF تميل للتمييز عبر المجموعات جميعها، ولكنها بشكل خاص تميز مجموعة الصعوبات الشديدة عن المجموعات الأخرى.

وكما هو الحال لدالة التمييز الواحدة، فإن دوال التمييز المتعددة يمكن استخدامها لتصنيف الأفراد إلى مجموعات على أساس درجاتها على دوال التمييز. والخطوة الأولى تكمن في استخدام العينة لتقدير دوال التمييز ومتوسط كل مجموعة على كل دالة من دوال التمييز. ولتصنيف مفحوص جديد تحسب درجات دالة التمييز للمفحوص ثم نحسب المسافة بين هذه الدرجات، وكذلك يحسب مركز كل مجموعة، ثم يصنف المفحوص كواحد من أعضاء المجموعة التي يكون مركزها أقرب لدرجات دالة تمييز المفحوص نفسه.



شكل (11-1): شكل المجموعات في الفضاء التمييزي.

الخلاصة:

وصف هذا الفصل طرائق إحصائية في دراسة الصدق والتي تستخدم اثنين أو أكثر من المتنبآت. ويمكن استخدام معامل الارتباط الجزئي عندما يكون مستخدم الاختبار مهتماً بدراسة العلاقة بين المحك والمتنبئ الثاني. وبافتراض تجانس المفحوصين على المتنبئ الأول فإن حساب هذا الإحصائي وتفسيره تم توضيحه للحالة التي تستخدم متنبأين اثنين، كذلك عُرِضَ بصيغة رياضية لتصحيح معامل الارتباط الجزئي من أثر أخطاء القياس.

وعند إجراء دراسة الصدق لتحديد الأثر الموزون لمتنبأين اثنين أو أكثر فإن الطريقة الإحصائية المناسبة هي تحليل الانحدار المتعدد. وعرض مثال مع حسابات توضيحية لحساب معاملات الانحدار والقاطع في معادلة التنبؤ الذي يستخدم متنبأين اثنين، بالإضافة إلى الصيغة العامة لمعادلة التنبؤ التي تستخدم أكثر من متنبأين اثنين. ولوحظ أن حساب معاملات الانحدار في هذه الحالة معقدة بدرجة أكبر وتطبق عادة باستخدام الحاسوب. ويتم اختيار قيم معاملات الانحدار لمتغيري المتنبئ والقاطع بحيث يتم خفض الفجوة بن درجات المحك الفعلية ودرجات المحك المتنبأ بها للعينة المستخدمة في دراسة الصدق. وقد تم وصف وتوضيح طريقة لتحديد أفضل مجموعة فرعية من المتنبآت. ويطلق على الأسلوب المستخدم هنا اسم مجموعات الانحدار الفرعية المحتملة جميعها.

وعندما يكون مستخدم الاختبار مهتماً بدقة التنبؤ من معادلة الانحدار فإن إحصائيات كل من مربع معامل الارتباط ومربع معامل الارتباط المتعدد للصدق تعد إحصائيات مفيدة. فالأول يحدد على أنه الارتباط بين أداءات المفحوصين المحكية (y) والتنبؤ بأدائهم المحكي (y') باستخدام معادلة تنبؤ المجتمع. والأخير هو الارتباط بين أداءات المفحوصين المحكية والتنبؤ بأداءاتهم المحكية (y') باستخدام معادلة تنبؤ العينة. وتم عرض توضيح لإثبات كيف أن هذين الارتباطين قد يختلفا للبيانات نفسها عندما تكون نسبة حجم العينة إلى عدد المتنبآت قليلاً. وحيث أن التنبؤ بأداء المفحوصين يجب إجراؤه على أساس بيانات العينة فإن مربع معامل الارتباط المتعدد للصدق التقاطعي هو الأكثر أهمية في تفسير نتائج الصدق.

وعندما يكون مستخدم الاختبار مهتماً بدرجة الدقة في التنبؤ بدرجات

المفحوصين المنفردين فإنه يوجد إحصائيين اثنين مهمين هما الخطأ المعياري للتقدير والخطأ المعياري لـ (y^1)، ويفضل الخطأ المعياري لـ (y^1) على الآخر لأنه يأخذ بعين الاعتبار حجم العينة ودرجة العلاقة بين المتنبآت. ويفترض أن للخطأ المعياري لـ (y^1) قيماً مختلفة للقيم المختلفة لمتغيرات المتنبئ ويكون الأقل للقيم القريبة من المتوسط.

وأخيراً وصفت استخدامات تحليل التمييز في المواقف التي يأمل مستخدم الاختبار بوساطتها استخدام متغيرات المتنبئ في تصنيف المفحوصين إلى فئات مختلفة على المحك المختار. وفي إجراء دراسة الصدق مثل هذه يمكن استخلاص الخطوات الآتية:

1. تحديد معادلة تمييز وظيفية والتي هي جمع موزون لمتغيرات المتنبئ.
2. تحديد درجة القطع لدالة التمييز لاستخدامها كأساس للتصنيف.
3. تقييم دقة التصنيف الناجم عن استخدام دالة التمييز من خلال احتمالية التصنيف الخاطئ.

التمارين:

1/ استخدم اختبار مينسوتا لتقييم القراءة (MRA) لتقييم مهارات القراءة لطلبة كليات المجتمع، وكلية إدارة الأعمال والمعاهد المهنية التربوية الثانوية وما بعد الثانوية، بينما صمم اختبار ستانفورد للمهارات الأكاديمية (TASK) للطلبة في الصفوف من الثامن إلى الثالث عشر. وقد طبق كل من براون وشانغ (1982) اختبارات القراءة الفرعية MRA و TASK على (50) من الطلبة المنخرطين في المعاهد المهنية التقنية. وقد استخدم هؤلاء متوسطات الطلبة وحسبوا مصفوفة الارتباطات لهذه المتغيرات الثلاثة:

GPA	TASK	MAR	
0.60	0.80	1.00	MRA
0.67	1.00		TASK
1.00			GPA

احسب وفسر الارتباط الجزئي بين MRA و GPA بضبط TASK.

2/ في المثال المتعلق بدراسة هارتلي وهارتلي (1976) كانت نسبة حجم العينة إلى المتنبئ 2.59: 1، وأقل بكثير من القانون التجريبي المقترح وهو 10: 1، وقيمة R^2 المدونة = 0.86 و $R_{cv}{}^2$ المدونة 0.64، ويقترح هذا المثال أن النسبة القليلة لحجم العينة للمتنبئ لها أثر على الصدق التقاطعي. احسب $R_{cv}{}^2$ المناظرة لـ R^2=0.50 و R^2=0.70 لموقف عدد مفحوصين (57) ومتنبآته (22) (أي عدد المفحوصين والمتنبآت لدراسة هارتلي وهارتلي/) ماذا يمكن أن تستنتج عن أنواع المواقف التي يكون فيها نسبة المفحوصين إلى المتنبئين من 3 إلى 1 مزعجاً بدلالة الصدق التقاطعي.

3/ دوّن تشيسوم وهوينز (Chissom & Hoenes, 1976) المتوسطات والانحرافات المعيارية والارتباطات لاختباري ذكاء متحررين من الثقافة: اختبار D-48واختبار الذكاء المتحرر من الثقافة (CFIT) والاختبار التحصيلي لرابطة الأبحاث العلمية SRA على (150) طالباً في الصف الثامن. وكانت مصفوفة الارتباطات والمتوسطات والانحرافات المعيارية على النحو الآتي:

SRA	CFIT	D-48	
0.61	0.69	1.00	D-48
0.66	1.00		CFIT
1.00			SRA
182.67	13.62	23.12	μ
58.41	7.66	7.88	σ

أ- أسس معادلة انحدار التنبؤ بـ SRA من D-48 و CFIT.

ب- ما درجة SRA المتنبأ بها لطالب درجته 30 عل D - 48 و 20 علىCFIT.

ج- احسب الخطأ المعياري للتقدير.

د- في أي مدى من الدرجات يمكن أن تكون نسبة الثقة 68% لدرجة SRA لطالب درجته 30 على D - 48 و 20 على CFIT (يمكنك استخدام الخطأ المعياري للتقدير في الإجابة عن هذا السؤال).

هـ- باستخدام البيانات المتوافرة، هل أن الخطأ المعياري للتقدير مُغالى فيه بدلالة دقة التنبؤ المعتمد على معادلة انحدار العينة؟ لماذا؟.

4/ اهتم في مصحة معالجة من الإدمان الكحولي بتطوير طريقة للتنبؤ بأي المرضى سيعود

للشرب غير المضبوط خلال مدة ستة أشهر بعد خروجه من المصحة. ومن خلال مقابلات أجريت بعد ستة أشهر من الخروج وجد أن (10) مرضى كانوا مضبوطين في الشراب وحدد (10) مرضى على انهم شاربون غير مضبوطين. وتوافرت بيانات على (5) متغيرات قيست أثناء المقابلة الابتدائية. وهذه المتغيرات هي: اختبار متشغان لغربلة الشاربين(X_1)، والدخل السنوي بآلاف الدولارات (X_2)، والدرجات على مقياس مركز الضبط الداخلي - الخارجي لروتر (X_3)، والتعليم (X_4)، والعمر عند أول مرة شرب (X_5)وفيما يأتي البيانات المستخدمة في حساب دالة التمييز:

المجموعة	المتغيرات				
	X1	X2	X3	X4	X5
الشاربين المضبوطين	17	24	6	16	8
	14	21	11	11	12
	16	23	8	14	17
	18	26	4	12	12
	18	25	10	9	16
	17	22	12	17	14
	15	23	6	9	17
	19	25	11	10	15
	17	24	8	16	14
	18	22	5	11	16
الشاييين غير المضبوطين	14	25	5	11	15
	17	28	7	13	14
	18	27	7	17	13
	12	24	4	13	11
	17	26	8	16	16
	18	29	6	15	13
	11	22	6	17	12
	14	25	9	16	15
	17	28	5	14	15
	18	26	8	16	14

ونتج عن هذه البيانات دالة التمييز الآتية:

$$Y= -.34 X_1 + .40 X_2 - 0.01 X_5 + .14 X_4 + 0.09 X 5$$

أ- استخدم هذه المعادلة في تطوير طريقة لتصنيف المرضى مستقبلاً.

ب- كمشرف على مصحة معالجة مدمني الكحول كيف تستخدم هذه المعادلة.

ج- احسب دقة هذا القانون بتطبيقه على بيانات مناسبة.

5/ يتألف مقياس تقدير السلوك الأولي لديڤروكس (Devereux) (DBRS)من (47) فقرة صممت لتمثل المدى الكامل لسلوكات الطلبة التي تظهر في الصفوف المنتظمة. وينتج عن تصحيح المقياس (11) درجة اختبار فرعية، والدرجات العالية في المقاييس جميعها (عدا الاستيعاب والمبادرات الإبداعية). تمثل سلوك اقل ايجابية. وتم قياس (217) طالباً في الصف السادس (تضمنوا طلبة التخصص الرئيسي بالإضافة إلى (184) طالباً منتظماً) من قبل المعلمين على مقياس تقدير السلوك الأولي (DBRS). وكان متوسط طلبة التخصص الرئيسي أعلى من متوسط الطلبة المتضمنين على الاختبارات الفرعية جميعها عدا الاستيعاب، والفروق الأقل كانت على اختبار الحاجة إلى الحميمية واختبار المبادرات الإبداعية. وحللت البيانات باستخدام تحليل التمييز، وفيما يأتي أوزانا لتمييز المعيارية (SDW) ومعاملات الارتباط التمييزي. اكتب تفسيراً مختصراً لتغير التمييز، وماذا تقترح هذا فيما يتعلق بإدراك المعلمين للفوارق بين طلبة التخصص الرئيسي والطلبة المنتظمين.

الاختبار الفرعي	SPW	الارتباط
الإزعاج الصفي	0.25	0.76
قلة الصبر	-0.35	0.67
عدم الاحترام – التحدي	0.09	0.71
اللوم الخارجي	0.24	0.77
قلق التحصيل	0.12	0.54
الاعتماد الخارجي	-0.29	0.65
الاستيعاب	-0.78	-0.64
المهمل – المنسحب	0.10	0.67
الاستجابات غير المناسبة	0.48	0.85
المبادرات الإبداعية	0.40	0.22
الحاجة إلى الحميمية	-0.07	0.13

الفصل الثاني عشر

12

التحيز في الاختيار

الفصل الثاني عشر

التحيز في الاختيار

ظهرت في أواخر الستينات قضية استخدام الاختبارات في الاختيار للوظيفة وفي المؤسسات التربوية. وبدأ يظهر الاهتمام المتزايد في الدراسات السيكومترية، ومع أن قضية تحيز درجات الاختبار شغلت بال مطوري الاختبارات ومستخدميها، إلا أنه لا توجد صياغة واضحة لتعريف تحيز الاختبار يمكن تطبيقها بموضوعية للكشف عن تأثيره أوحذف هذا التأثير في مواقف الاختبار، وفي البداية كانت المحاولات لتطوير تعريف تحيز الاختبار من خلال خطوط الانحدار وذلك للتنبؤ عن الأداء بوساطة درجات الاختبار في مجموعات المفحوصين الرئيسية والفرعية (لمراجعة استخدام خطوط الانحدار انظر الفصلين الثاني والحادي عشر)، وللاهتمام بمثل هذه الأهداف فإن التعريف الاحصائي لتحيز الاختبار يأمل بناء اختبارات يحذف منها التحيز في عملية الاختيار، ومع ذلك فإن خط الهجوم هذا انخفض فوراً في وجه النمو المتزايد لقضية التحيز في الاختيار، وعدت على أنها مسألة قيم سياسية اجتماعية، وبغض النظر عن طريقة مقارنة خطوط الانحدار بين المجموعات الرئيسة والفرعية فإن طريقة الاختيار أو قانونه يمكن اقتراحها لتحقق أية أهداف سياسية اجتماعية بالمقابل، اقترحت العديد من طرائق الاختيار لتعكس مواقف قيمية مختلفة حول عدالة الاختيار.

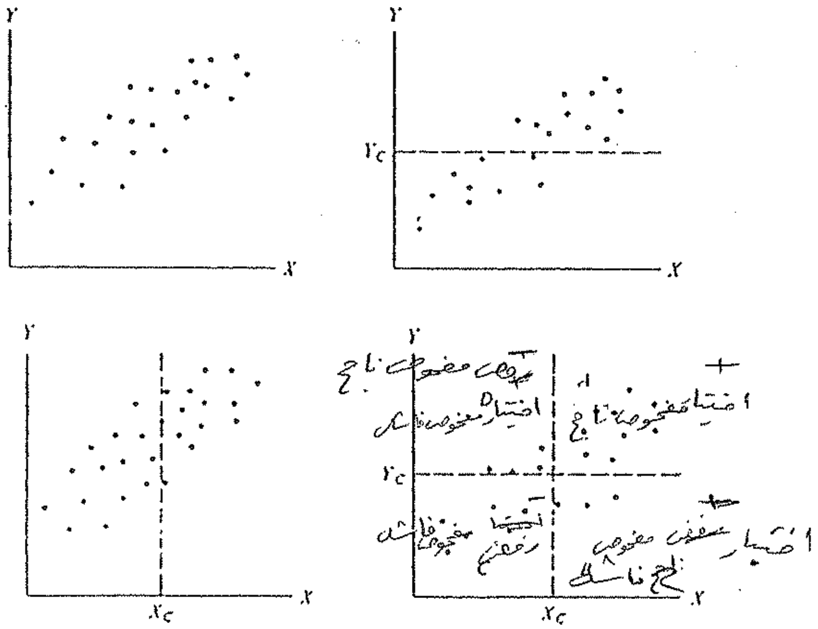
وبتتبع هذا التطور فإن العديد من المؤلفين:

(Cronbach, 1976, Gross & su, 1975, Petersons Novick, 1976)

أشاروا إلى ملائمة نظرية القرار لتطوير طرائق العدالة في الاختيار، ويتوافر الآن اتفاق أساسي على أن نظرية القرار تزودنا بأسلوب نسبي وتقني. وكنتيجة، فإن العديد من طرائق الاختيار المبكرة التي كانت صياغتها على غير طريقة نظرية القرار لم تستمر تطبيقاتها طويلاً، ومع ذلك فإن فهم الكتابات الحالية في القياس المتعلقة بالتحيز في الاختيار تتطلب معرفة بعض المحاولات المبكرة لتعريف التحيز ولتطوير نماذج اختيار تسبق تلك التي اعتمدت نظرية القرار، كذلك إضافة إلى الحديث عن تاريخ هذه التعريفات وطرائق الاختيار، فإن هذا الفصل يقدم مدخلاً إلى نظرية القرار واستخداماتها في مشكلات الاختيار، وقبل البدء بأي من هذه الموضوعات يجب علينا أن نقدم تعريفاً لبعض المصطلحات والمفاهيم الأساسية.

مفاهيم ومصطلحات أساسية:

في البدء، دعنا نركز على اختيار مفحوصين من مجموعة واحدة، ويوضح شكل (12-1أ) تمثيلاً بيانياً اعتماداً على درجات الاختبار X ودرجات المحك Y، وهذا مشابه لما نحصل عليه من دراسة الصدق البسيطة، افترض أنه اختيرت نقطة على المحور Y بحيث تقسم المفحوصين إلى مجموعتين (ناجحة وراسبة) وكما هو موضح في الرسم البياني في الشكل (12 - 1ب)، افترض أيضاً أنه تم اختيار قطع على محور السينات بحيث يمكن اختيار المفحوص للبرنامج إذا كانت درجته على درجة القطع أو أعلى منها، أما إن كانت درجته دون هذه الدرجة فإنه سيرفض في هذا البرنامج (شكل 12 -1ج)، والآن فإنه لأي مفحوص قد تكون نتيجتان محتملتان، أما نجاح أو فشل في المحك، كذلك فإن هنالك قراران محتملان بالاعتماد على المتنبيء: القبول أو الرفض، وينتج عن الارتباطات الممكنة بين النتاجين وبديلي القرار أربعة احداث ممكنة تمثلها الرباعيات المبنية في الرسم البياني في الشكل (12 - 1د)، ويتضمن الربع الأول الذي يرمز له بالرمز A إحداث حقيقية موجبة (أي اختيار المفحوص الناجح) ويتضمن الربع الثاني المرمز له B احداث خاطئة موجبة (أي اختيار المفحوص الفاشل)، ويتضمن الربع الثالث المرمز له C احداث حقيقية سالبة (أي رفض المفحوص الفاشل)، ويتضمن الربع الرابع المرمز له D إحداث أحداث خاطئة سالبة (أي رفض المفحوص الناجح).



شكل (12-1): تمثيل بياني توضيحي بنقاط تعتمد على درجات الاختبار (X) ودرجات المحك (Y).

وفي التاريخ الطويل للقياس النفسي تستخدم ثلاثة مصطلحات تعتمد على التكرارات أو عدد الحالات التي تقع في كل ربع في الشكل (12 –1د)، وهذه هي:

1/ النسبة القاعدية: وهي نسبة المفحوصين الذين يمكن أن ينجحوا في المحك، وتحسب النسبة من العلاقة:

$$\frac{D + A}{D + C + B + A}$$ ، حيث تمثل الرموز عدد المفحوصين في الأرباع الأربعة.

2/ نسبة الاختيار: وهي نسبة المفحوصين الذي سيختاروا لملء الشواغر الوظيفية، ويحسبوا باستخدام العلاقة

$$\frac{B + A}{D + C + B + A}$$

3/ نسبة النجاح: وهي نسبة المفحوصين الذين نجحوا، وتحسب من العلاقة $\frac{A}{B + A}$

وقد طور كل من تايلور وراشل **(Taylor & Russel, 1939)** مجموعة جداول مكثفة تبين تفاعل الصدق ونسبة الاختيار التي تؤثر في نسبة النجاح، وعلى القارئ ملاحظة أن الاختيار العشوائي للمفحوصين من مجموع المتقدمين يؤدي إلى أن تكون نسبة النجاح مساوية للنسبة القاعدية، لهذا فإن مقارنة نسبة النجاح الناتجة عن استخدام متنبيء معين بالنسبة القاعدية تشير إلى مدى تطور قرار الاختيار باستخدام الاختبار بالنسبة لنوعية القرارات المتخذة بالاختيار العشوائي للمفحوصين.

ويبين الجدول (12–1) نسبة النجاح لمعاملات ارتباط مختلفة بين الصدق التنبؤي(ρ_{xy}) ونسبة الاختيار. وقد حسبت نسب النجاح عند نجاح 50 % من المفحوصين من المجموع الكلي على المحك (على سبيل المثال عندما تكون النسبة القاعدية 50 %)، وبتفحص العمود عند ρ_{xy} = صفر، نرى أن نسبة النجاح 50 %، وعلى هذا فإن استخدام الاختبار للاختيار ينتج عنه نسبة نجاح مساوية للنسبة القاعدية، بمعنى أن استخدام اختبار له معامل ارتباط = ρ_{xy} صفر لا يكون له أثر أكثر مما هو للاختيار العشوائي للمستخدمين، وبفحص العمود عند ($\rho_{xy} = 0.2$)، نجد أنه عند نسبة اختيار 0.10 وتكون نسبة النجاح 0.64 ، لاحظ أنه مع نسب الاختيار المنخفضة (0.10) واختبار معامل صدقه منخفض ($\rho_{xy} = 0.2$) يمكن أن ينتج عنه نسبة نجاح أعلى من النسبة القاعدية الناتجة عن استراتيجية الاختيار العشوائي، كذلك إذا كانت نسبة الاختيار 0.70 فإن نسبة النجاح 0.54، لذلك فإن اجبرنا على

اختيار نسبة أساسية من المتقدمين فإن نسبة النجاح لا يمكن أن تكون أكبر من النسبة القاعدية، ولا تظهر هذه الظاهرة فقط عندما يكون معامل الصدق منخفضاً، ولكن مع اختبارات لها معامل صدق عالٍ، وبشكل أساسي إن كان $\rho_{xy} = 0.80$ ونسبة الاختيار 0.90 فإن نسبة النجاح تكون 0.55 فقط، وفي الحالة المتطرفة إن أجبرنا على استخدام نسبة اختيار 100 % (أي اختيار المتقدمين جميعهم) فإن نسبة النجاح تهبط إلى 50 %، لهذا فإنه لا ميزة لاختيار اختبار ذي صدق عالٍ لهذا الموقف، ويمكننا تلخيص العلاقة بين هذه المفاهيم على النحو الآتي، وعلى افتراض أن النسبة القاعدية تبقى ثابتة لمجموعة المفحوصين:

1- تكون نسبة النجاح عالية عندما يكون معامل الصدق عالياً ونسبة الاختيار منخفضة.

2- عند ازدياد نسبة الاختيار ودون تغيير معامل الصدق فإن نسبة النجاح تقل.

3- بثبات نسبة الاختيار وزيادة معامل الصدق فإن نسبة النجاح تزداد.

ومن المهم ملاحظة أن كل جدول من جداول تايلور ورشل تم بناءه على أساس نسبة قاعدية مختلفة، وباختلاف النسب القاعدية (وكما هي للمجموعة الفرعية المختلفة) فإنه يلزم معاملات صدق مختلفة ونسبة اختيار مختلفة لتحقيق نسبة النجاح نفسها، على سبيل المثال: إذا زادت النسبة القاعدية فإننا نحتاج لمعامل صدق أقل للحصول على نسبة النجاح نفسها (وذلك عند نسبة اختيار معينة)، بالمقابل إذا انخفضت النسبة القاعدية ولم يتغير صدق الاختبار فإنه يلزم نسبة اختيار أقل للحفاظ على نسبة النجاح نفسها، ومع ذلك لم يستمر استخدام جداول تايلور ورشل كثيراً في التنبؤ بنسب النجاح في مواقف معينة، ومن المهم لمستخدمي.

الجدول (12 –1): نسب النجاح لتجمعات مختلفة لمعاملات الصدق ونسب الاختيار عند نسبة نجاح قاعدية 0.50 .

معامل الصدق $\hat{\rho}_{xy}$					نسبة الاختيار
0.80	**0.60**	**0.40**	**0.20**	**0.00**	
0.99	0.90	0.78	0.64	0.50	0.10
0.90	0.79	0.69	0.59	0.50	0.30
0.80	0.70	0.63	0.56	0.50	0.50
0.67	0.62	0.58	0.54	0.50	0.70
0.55	0.54	0.53	0.52	0.50	0.90
0.50	0.50	0.50	0.50	0.50	1.00

مقتبس عن:

H.C. Taylor and J.T. Russell The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied psychology, 23,565-578.

الاختبارات أن يميزوا (حقيقة الحياة) احدى السيكومتريات التي يتبنوها، وتحديداً فإنه لمجموعتين من المفحوصين ولنسب قاعدية مختلفة تم اختيارهم بنسب الاختيار نفسها وباستخدام اختبار له معامل الصدق نفسه لكلا المجموعتين، كذلك فلمجموعتي مفحوصين وبنسب قاعدية مختلفة ونريد تحقيق نسب النجاح نفسها لكليهما، فإن هذا غير ممكن باستخدام اختبارات لها معاملات صدق ونسب اختيار متماثلة، ويعد تمييز هذه المبادئ متطلباً سابقاً لفهم أعمق للأدب النفسي المتعلق بالتحيز الثقافي ونماذج الاختيار العادل.

المجموعات الرئيسية والمجموعات الفرعية:

من الضروري تكوين معنى لمصطلحات المجموعات الرئيسية والفرعية، كما تستخدم في السياقات النفسية، وتشير المجموعة الرئيسية إلى المجموعة التي تتضمن العدد الأكبر من المتقدمين إلى مركز استخدام أو مؤسسة تربوية معينة، وتشير المجموعة الثانوية إلى المجموعة التي تحوي أقل عدد من المرشحين من مجموع المتقدمين، فعلى سبيل المثال إذا نظرنا إلى المتقدمين للقبول في مدرسة التمريض فإننا نجد أن الإناث يمثلن المجموعة الرئيسة، بينما يمثل الذكور المجموعة الثانوية، وإذا نظرنا إلى المتقدمين إلى كلية الطب البيطري فإن الموقف ينعكس إذ يمثل الذكور المجموعة الرئيسة وتمثل الإناث المجموعة الثانوية، وتتحدد المجموعات الرئيسية والثانوية بوساطة الخصائص الديموغرافية التي تهم مستخدم الاختبار (مثل العنصر أو العرق، الخلفية الثقافية، العمر، العجز الطبيعي، .... الخ).

إن قضية عدالة الاختيار تظهر حتى عندما تكون المجموعة الرئيسة والثانوية متساوية في العدد، افترض أن جامعة تحوي العديد من كليات المجتمع وتسمح بتحويل التخصص في السنة الثالثة، وهنا تظهر قضية مهمة وهي العدالة في اختيار المجموعات التي تدرس في الجامعة في السنتين الأولى والثانية، ومجموعة الطلبة المحولين في السنة الثالثة حتى عندما تكون المجموعتان ممثلتين بالتساوي من مجموعة المتقدمين، ففي مثل هذه الحالة ليس المهم كيفية تصنيف المجموعتين.

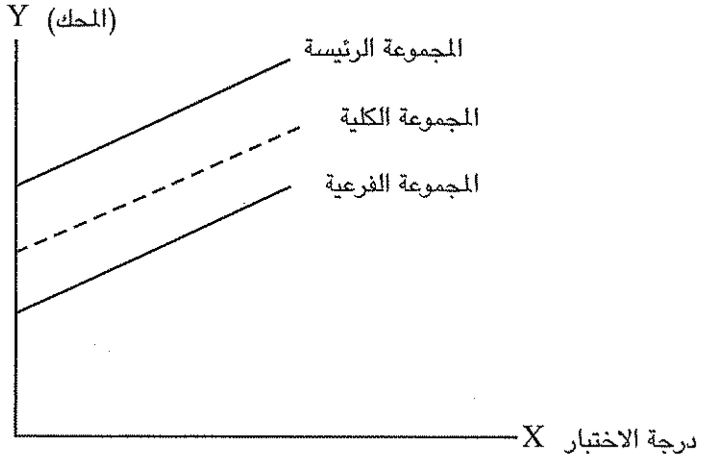
التعريف النفسي للتحيز:

أشار كليري (Cleary, 1968) إلى أن الاختبار يكون متحيزاً لإحدى المجموعات الفرعية في المجتمع إذا ظهرت اخطاء غير صفرية في التنبؤ المتعلق بالأعضاء، ويوضح الشكل (12-2)

أحد المواقف التي يمثلها كليري، ويكون خط الانحدار هنا للمجموعة الكلية أعلى من خط الانحدار للمجموعة الثانوية وادنى من خط انحدار المجموعة الرئيسية.

وفي المثال يكون استخدام خط الانحدار للتنبؤ بالأداء غير متنبأ به للمجموعة الرئيسية (Un predicted)ويكون أكثر من متنبأ به (Over predicted) للمجموعة الثانوية، لهذا فإن استخدام خط الانحدار للمجموعة الكلية يؤدي إلى تنبؤ متحيز، وفسّر كل من دارلنغتون (Dar-lington, 1971) وهنتر وشميدت (Hunter & Schmidt, 1976) ولين (Linn, 1973) تعريف كليري بأن الاختبار لا يكون متحيزاً إذا كانت خطوط انحدار المجموعة الرئيسة والمجموعة الثانوية هي نفسها، فإن كانت خطوط الانحدار هي نفسها لمجموعة مفحوصين متجانسين على المتنبئ فإنه لا توجد فروق بين درجات المحك المتنبئ لكل من المجموعة الرئيسية والثانوية، لهذا فإن تعريف كليري وكما فسره دارلنغتون والآخرون يتطلب تساوي قدرة الطلبة كما تقاس بالمتنبئ لمن حصلوا على درجة المحك نفسها للاختبار غير المتحيز.

وعدل جنسن (Jenssen, 1980) تعريف الانحدار المتساوي لتحيز الاختبار، وقد برهن على أن الاختبار لا يكون متحيزاً إذا:

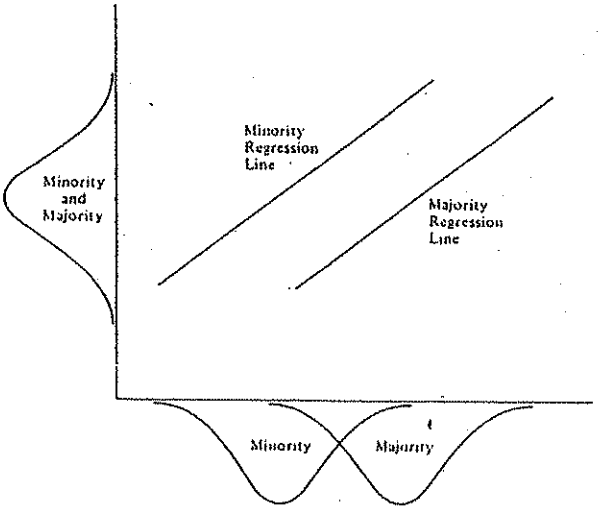


شكل (2–12):

خطوط الانحدار للمجموعات الرئيسية والثانوية التي توضح تعريف كليري لتحيز الاختبار

1. كانت خطوط الانحدار لدرجات المحك على درجات المتنبئ الحقيقية هي نفسها لكلا المجموعتين و (2) كانت الأخطاء المعيارية لقيم المتنبئ (انظر فصل 11) هي نفسها لكلا المجموعتين، ويدعم الشرط الأول بحقيقة أن خطوط الانحدار على الدرجات الملاحظة للمتنبئ تكون غير متساوية، وببساطة بسبب أخطاء القياس حتى إن كانت خطوط انحدار درجات المتنبيء الحقيقية متساوية، والشرط الثاني يعني ببساطة أن المتنبئين لهما درجة الدقة نفسها لكلا المجموعتين.

وفسر مكنمار (McNemar, 1975) تعريف كليري على أن طريقة الاختيار تكون غير متحيزة إذا استخدمت خطوط انحدار منفصلة لكلا المجموعتين الرئيسية والثانوية وكانت هذه الخطوط غير متماثلة، إن استخدام خطوط انحدار منفصلة يلغي مشكلة اتساق الأخطاء غير الصفرية وبالتالي تنبؤ غير متحيز، لاحظ الانتقال في موقف مكنمار من الاهتمام بالاختبار غير المتحيز إلى طريقة اختيار غير متحيزة، وهذه الازاحة توضح كيفية تحول السؤال عن تحيز الاختبار إلى سؤال عن تحيز الاختيار، وعلى القارئ أن يميز أن تبني موقف مكنمار يعني أنّ المفحوص ذو الدرجة الأعلى على المتنبئ يجب قبوله بغض النظر عن المجموعة التي ينتمي إليها، لهذا فإن تعريف مكنمار لطريقة الاختيار العادل تتضمن عدالة متساوية في اختيار المفحوصين الذين تتنبأ لهم بمقدرات أعلى، ومن الواضح أن هذا موقع ذو قيمة محمولة (Value-laden position) بالمقابل، وعلى سبيل المثال: يكون الاختيار عادلاً عندما يختار المفحوص ذو الدرجة الأعلى لمركز ما، وعلى النقيض فإن مركز مكنمار لا يؤدي بالضرورة لاختيار المتقدمين ذوي الدرجات الأعلى على المتنبئ.



شكل (12-3): خطوط الانحدار الثانوية والرئيسة مع الخط الثانوي يعلو الخط الرئيسي

افترض على سبيل المثال في الشكل (12–3) كان توزيع درجات المحك نفسه للمجموعة الرئيسة والمجموعة الثانوية، ولكن أداء المجموعة الرئيسية على المتنبئ كان أفضل من المجموعة الثانوية، ففي هذه الحالة تكون درجة المفحوص على المتنبئ من المجموعة الثانوية أقل من درجة المفحوص من المجموعة الرئيسة، ولكن درجته على المحك أعلى لذا سيتم اختياره بدلاً من مفحوص المجموعة الرئيسية.

وقدم ثورنداك (Thorndike, 1971) تعريفاً ثانياً للتحيز مؤكداً على أن الاختبار إذا كان لأي درجة مناسبة على المحك (مثل درجة النجاح على المحك) وبدرجة قطع واحدة على المتنبئ يمكن أن توجد بنسب نجاح لكلا المجموعتين الرئيسية والثانوية نفسها على المحك وهذا يعني أن الاختبار عادل اساساً إذا كانت المجموعات الرئيسة والثانوية تبعد عن المتنبئ كما هو على المحك، ولتوضيح هذا التعريف بدرجة أكبر افترض أن الانحراف المعياري على المتنبئ كان نفسه لكلا المجموعتين، وهذا ينطبق أيضاً على المحك وعندما يتوافق المتنبئ مع تعريف ثورنديك فإن معامل ارتباط بوينت بايسيريال بين المجموعة الرئيسية أو الثانوية مع درجة المتنبئ ستساوي معامل ارتباط بونيت بايسيريال بين المجموعة ودرجة المحك.

وأضاف دارلنغتون (Darlington, 1971) تعريفين إضافيين للتحيز، ولكنه لم يفضل أحدهما على الآخر أحد التعريفين هو أن الاختبار يكون غير متحيز إذا كان توزيع الدرجات على المتنبئ هو نفسه لكلا المجموعتين، والتعريف الأخر، هو أن الاختبار يكون غير متحيز إذا كان انحدار المتنبئ على المحك هو نفسه لكلا المجموعتين، وكما لاحظنا سابقاً، وبعد اقتراح هذه التعريفات لتحيز الاختيار يتضاءل الاهتمام بتعريف التحيز من خلال العلاقة بين المتنبئ والمحك فضلاً عن التوجه إلى مشكلة التعريف من خلال العدالة في الاختيار، وتطوير طرائق لاستخدام هذه التعريفات، وسنراجع بعض هذه الجهود في الجزء الآتي.

طرائق الاختيار العادلة:

من المهم عند البحث عن أسئلة العدالة في الاختيار التمييز بين نوعين من القرارات: الفردية والمؤسساتية ويظهر اتخاذ القرارات الفردية عندما يكون الهدف الأساسي لعملية اتخاذ القرار هو الزيادة القصوى للمنفعة الفردية، فعلى سبيل المثال عندما ينشيء المرشد قائمة الميول المهنية وينصح المفحوص باختيار المهنة الأنسب، فإن درجة الاختبار التي تحدد ميل المفحوص تستخدم هنا لاتخاذ قرار فردي بالمقابل، فإن اتخاذ قرار مؤسسي يظهر عندما

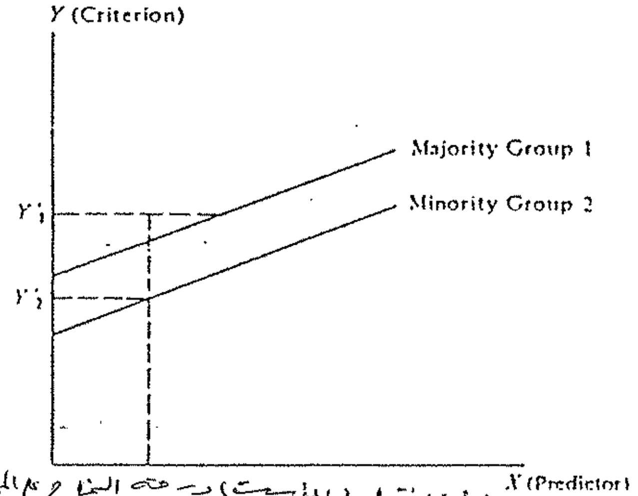
يكون الهدف الأساسي لاتخاذ القرار هو تحقيق أهداف خاصة بالمؤسسة (مثل اختيار المتقدمين الذين احتمال نجاحهم أكبر)، ودون الأخذ بعين الاعتبار الأفضل لأي من المتقدمين، النقطة الأولى التي سنكونها بالنماذج السيكومترية للعدالة ستأخذ بعين الاعتبار هو أن قرار الاختيار هو اتخاذ قرار مؤسسي، وعلى هذا فإن طرائق عدالة الاختيار ستأخذ بعين الاعتبار أهداف المؤسسة.

وقد تم اقتراح العديد من نماذج عدالة الاختيار في الأدب السيكومتري، وسيتم توضيح ستة نماذج في هذا الفصل، وهذه النماذج هي نموذج الانحدار، ونموذج المخاطرة المتساوية للمستخدمين، ونموذج دارلنغتون، ونموذج الاحتمالات المتساوية، ونموذج النسب المتساوية، ونموذج الاحتمالات الشرطية، وقد لا تكون التفاصيل التقنية من اهتمامات القارئ، وقد كتب عن هذه التفصيلات كول (Cole, 1973) والذي يزودنا أيضاً بمقارنة هذه النماذج بعضها ببعض من خلال درجة القطع على المتنبئ ونسبة المفحوصين المقبولين من المجموعة الثانوية وخصائص أخرى.

نموذج الانحدار:

يعتمد نموذج الانحدار على افتراض أن الاختيار يكون عادلاً عندما يتم اختيار الذين درجاتهم على المحك أعلى بغض النظر عن المجموعة التي ينتمون إليها، لهذا فعند قبول (100) طالب للدراسة في تخصص معين يكون محك الاختيار هو اختيار الذين حصلوا على أعلى (100) درجة على المتنبيء. ولتحقيق هذا يتم مطابقة خطوط انحدار منفصلة لكلا المجموعتين، ويتم الاختيار من المتقدمين الذين حصلوا على أعلى درجات المتنبئ، وحتى عندما تتطابق خطوط الانحدار للمجموعتين فإن المفحوصين الذين لهم الدرجة نفسها على المتنبئ ومن مجموعات مختلفة فإن التنبؤ يكون بدرجات مختلفة على المحك (انظر شكل (12-4))، وكنتيجة فإن كان لمفحوصين اثنين الدرجة نفسها واحدهما في المجموعة الرئيسية والآخر في المجموعة الثانوية فإنه لا يتم اختيار الاثنين، لهذا فإن أي طريقة للاختيار غير هذه قد يكون لمن حصل على درجة أقل على المحك.

ولذلك، فللمؤسسة التي تهدف اختيار المتقدم المؤهل بدرجة أعلى من خلال الأداء على المحك بغض النظر عن المجموعة فإن هذه الطريقة تعد مناسبة كنموذج اختيار.



شكل (12–4) استخدام نموذج الانحدار لتحديد درجة قطع المتنبئ للمجموعات الرئيسية والثانوية.

نموذج تكافؤ المخاطرة:

يتطلب هذا النموذج أن يحدد مستخدم الاختبار درجة النجاح على المحك وتحديد أقصى احتمالية ممكنة (أو مخاطرة) لفشل المقبول دون هذه الدرجة، ويتم هنا تحديد خطوط انحدار كلا المجموعتين ومن ثم حساب درجة التنبؤ على المحك (باستخدام معادلة التنبؤ المذكورة في الفصل الثاني) عند فشل المخاطرة، وفقط للمفحوصين الذين لهم فشل مخاطرة أقل من القيمة المحتملة هم الذين يتم اختيارهم، ويعتمد حساب احتمالية الفشل على افتراض التوزيع الطبيعي لدرجات المحك، وهو أنه عبر المفحوصين الذين لهم درجة معينة على المتنبّيء يفترض أن تتوزع درجاتهم توزيعاً طبيعياً على المحك، ويفترض أيضاً أن يكون تباينات التوزيع الاحتمالي الشرطي نفسه لدرجات المتنبئ جميعها، ومع هذه الافتراضات فإن احتمال الفشل يكون في المساحة تحت $Z = \frac{Y' - Y_p}{\sigma_{Xy}}$ في التوزيع الطبيعي المعياري، وفي هذه الصيغة تمثل y_p درجة النجاح على المحك، Y' هي الدرجة المتنبأة للمفحوص و σ_{Xy} هي الانحراف المعياري لدرجات المحك للمفحوصين جميعهم عند قيمة X المُعطاة، وهذا الانحراف المعياري يساوي الخطأ المعياري للتقدير، ويتم اختيار المتقدمين إذا كان احتمال فشلهم اقل من أقصى احتمالية للمخاطرة. لتوضيح ذلك: افترض أن درجة النجاح على المحك لفرد ما تقع عند 18، فإن اقصى مخاطرة = 0.35 والخطأ المعياري للتقدير للمجموعة الثانوية =8 وللمجموعة

الرئيسية = 4. وللمجموعة الثانوية فإن $Z = \frac{20-18}{8} = -0.25$ ، وللمجموعة الرئيسية

$Z = \frac{20-18}{8} = -0.5$، واحتمالية الفشل للمجموعة الثانوية = 0.40 وللرئيسة = 0.31.

ونتيجة لذلك فانه سيتم اختيار الفرد من المجموعة الرئيسية فقط أن نموذج المخاطرة المتكافئة يعمل جيداً إذا كانت المؤسسة تهدف ضبط الاحتمالية لفشل المستخدم، وتهدف لهذا بغض النظر عن خليط الأعضاء الناتج (ومن الجدير بالذكر هنا أنه في حالة تساوي الأخطاء المعيارية لكلا المجموعتين فإن نماذج المخاطرة المتكافئة والانحدار تؤدي إلى نتائج متشابهة، ويعد هذا النموذج قطعياً عندما تكون كلفة الاستخدام والتدريب أمراً أساسياً.

نموذج دارلنغتون:

أثبت دارلنغتون (Darlington, 1971) أن الطريقة المناسبة للاختيار العادل تكمن في تقرير كم هي القيمة لاختيار متقدم من المجموعة الثانوية، واستخدام هذه القيمة في اقتراح طريقة الاختيار، واقترح لتحديد هذه القيمة طرح السؤال "ما هو كم تناقض الدرجات على المحك بين أعضاء المجموعتين الرئيسية والثانوية، ويبقى اعتبار كل منهما مقبول وبالتساوي في عملية الاختيار"، فإن كانت قيمة التناقض K، فإن طريقة دارلنغتون في الاختيار تضيف هذا المقدار K لدرجة المحك المتنبأ بها لأفراد المجموعة الثانوية جميعهم، ويكون الأعضاء المختارين هم الذين لهم أعلى درجة على المحك بعد التعديل بإضافة k لأعضاء المجموعة الثانوية، وقد يكون مثل هذا النموذج مناسباً لمؤسسة ذات أولوية كبيرة في اختيار المتقدمين من المجموعة الثانوية أكثر من اختيارها لمن هم أكثر كفاءة أو ضبط احتمالية فشل المقبولين.

نموذج الاحتمالية المتساوية:

تتطلب الطرائق الثلاث السابقة تحديد خطوط انحدار منفصلة للمجموعات الرئيسية والثانوية ثم حساب درجات التنبؤ على المحك لكل متقدم باستخدام معامل انحدار مناسب لكل مجموعة، وتتطلب الطرائق الثلاث الآتية استخدام المفاهيم الصحيحة الموجبة، والخاطئة الموجبة، والصحيحة السالبة، والخاطئة السالبة وكما هو موضح في شكل (12-1د)، ويمكن فهم هذه الطرائق بشكل أفضل بافتراض أن مستخدم الاختبار ومن خلال عمليات التصديق السابقة حصل على خطوط بيانية تصف العلاقة بين درجة الاختبار والأداء على المحك لكلا المجموعتين الرئيسية والفرعية، وتفترض أيضاً أن مستخدم الاختبار حدد مستوى الحد الأدنى للقبول في الأداء على المحك والتي هي نفسها لكلا المجموعتين، وعلى مستخدم الاختبار في هذه الطرائق أن يحدد درجة القطع على المتنبيء لكل مجموعة، وبالتالي فإن

الخطوط البيانية لكلا المجموعتين تنقسم إلى أربعة أرباع (ربيعيات) والتي تتطابق مع شروط الاختيار الموصوفة في هذه النماذج.

يعتمد نموذج الاحتمالية المتساوية على أساس فلسفي يؤكد أنه لكل متقدم فرصة النجاح نفسها بغض النظر عن مجموعته، وبكلمات أخرى، فإن هذا يستلزم تساوي نسبة النجاح لكلا المجموعتين، وهذا يعني أنه إذا نجح 60% من المجموعة الرئيسية على المحك فإنه يجب أن ينجح 60% من المجموعة الثانوية على المحك، ولتحقيق هذا فإن درجة القطع على المتنبئ تحدد بأن تكون النسبة $\frac{A}{A+B}$ هي نفسها لكلا المجموعتين، حيث أن A هي الأحداث الصحيحة الموجبة و B هي الاحداث الخاطئة الموجبة (انظر شكل 12-1د)، وكنتيجة، فإن معالجة المتقدمين وبالدرجة نفسها على المتنبئ قد تتباين بالاعتماد على المجموعة التي ينتمون إليها، وبعد تحديدها فإن خصائص متساوية لكلا المجموعتين ستنجح، وقد وصف هذا النموذج من قبل لين (Linn, 1973) أولاً مع أنه لم يدافع عن استخدامه، وتكمن نقطة ضعف هذا النموذج في استحالة تطبيقه عندما يكون الاختلاف متطرف بين المجموعتين على درجات المتنبئ (Jensen, 1980)

نموذج النسبة الثابتة:

اعتماداً على القيم المتضمنة في تعريف ثورندايك لتحيز الاختبار اقترح نظام اختيار يكون عادلاً إذا تحددت درجة القطع على المتنبئ بحيث تكون نسبة الاختيار لكلا المجموعتين هي نفسها، وكما هي للنسبة القاعدية لكل منهم.

وعلى سبيل المثال افترض أن النسبة القاعدية للمجموعة الرئيسية 50% وللمجموعة الثانوية 50%، والنسبة القاعدية للمجموعة الثانوية هي نفسها (50% أيضاً)، وفي هذا النموذج عندما نختار 20% من المجموعة الرئيسية يجب أن نختار 20% من المجموعة الثانوية، ومثال آخر افترض أن النسبة القاعدية للمجموعة الرئيسية 60% وللمجموعة الثانوية 40%، لذا فإن النسبة بين النسبة القاعدية للمجموعة الثانوية هي $\frac{40}{60}$ اي ($\frac{2}{3}$) ولتتوافق هذه مع متطلبات ثورندايك فان نسبة المجموعة الثانوية الى الرئيسية في نسبة الاختيار يجب أن تكون 3:2، وعلى هذا فإذا اختير 20% من المجموعة الرئيسية فإن نسبة $20\% \times \frac{2}{3}$ =13.3 يجب أختيارها من المجموعة الثانوية.

ويمكن تبيان أن متطلبات نموذج ثورندايك يمكن تحقيقها عندما تكون النسبة $\frac{A+B}{A+D}$ هي نفسها لكلا المجموعتين (وثانية، يفضل عودة القارئ إلى الشكل (12-1د) إذ أنه يوضح

هذه القيم لكل من (A وB وC وD)، لهذا السبب فإن نموذج ثورندايك يسمى نموذج النسبة الثابتة، وكما هو الحال مع النماذج الأخرى فإن التطبيق يتطلب من مستخدم الاختبار تحديد درجات قطع مختلفة على الاختبار المتنبئ لمجموعات المفحوصين المختلفة.

نموذج الاحتمالية المشروط:

قدم هذا النموذج كول (Cole, 1973)، وركز فيه على مجموعة المتقدمين الذين سينجحون فعلاً، ويتطلب هذا النموذج احتمالية الاختيار نفسها لكل فرد يحتمل نجاحه بغض النظر عن المجموعة التي ينتسب إليها، وحتى يتحقق هذا، فإن النسبة $\frac{A}{D+A}$ يجب أن تكون نفسها للمجموعات جميعها (انظر شكل 12-1-د) , وثانية، لتحقيق هذا الهدف يجب تحديد درجات قطع مختلفة للمجموعات المختلفة، وتعرف النسبة $\frac{A}{D+A}$ على أنها احتمالية الاختيار المشروط بنجاح، وهذه تشير إلى اسم النموذج، ويطلق على هذا النموذج أيضاً اسم نموذج الفرص المتكافئة.

نقد النماذج:

نقد بترسون ونوفيك (Peterson & Novick, 1976) نماذج: النسبة الثابتة، والاحتمالية المتساوية والفرص المتكافئة من منطلقين، لتوضيح هذا، خذ على سبيل المثال نموذج ثورندايك الذي يتطلب أن تكون النسبة بين نسبة الاختيار إلى النسبة القاعدية ثابتة عبر المجموعات، وقد أثبت بترسون ونوفيك أن العدالة التي تتطلب نسبة ثابتة عبر المجموعات، وبالمنطق نفسه تتطلب العدالة أن تكون نسبة الرفض إلى نسبة فشل ثابتة عبر المجموعات (وهذا مكافئ لأن تكون نسبة الرفض لكلا المجموعتين مساوية لنسبة نسبة الفشل)، ويطلق على طريقة الاختيار التي تعتمد على هذا النموذج نموذج النسبة الثابتة المعاكس ووضح بترسون ونوفيك أن قرار الاختيار = ماعدا تحت ظروف خاصة = وفق نموذجي النسبة الثابتة والنسبة الثابتة المعاكس لا تكون هي نفسها، وكنتيجة يبدو أن كلا النموذجين غير مناسب، كذلك يمكن تقديم أدلة مشابهة بالنسبة لكل من نموذجي الاحتمالية المشروطة والاحتمالية المتكافئة.

وقد أشار كل من بترسون ونوفيك إلى أن كل فرد يجب أن يضع قرار شخصي حول فشل طريقة الاختيار في الاتفاق مع متطلبات نموذج ثورندايك ونموذج النسبة الثابتة المعاكس والتي تعد متناقضة منطقياً، وهذا يفتح الباب لاحياء وانتشار نموذج النسبة الثابتة كما هو لنماذج الاحتمالية المشروطة والاحتمالية المتكافئة، كذلك فإن لدى بترسون ونوفيك نقداً أكثر لهذه

النماذج، وتحديداً فهي غير مكررة للقرار النظري، ويظهر أن هذا النقد تم الموافقة عليه بصورة واسعة في أدبيات القياس، وفي الحقيقة فإن سوير وكول وكول (Sawyer, Cole & Cole, 1976) حاولوا ادخال القيم المتضمنة في نموذجي النسبة الثابتة والاحتمالية المشروطة في نموذج القرار النظري، وحيث أن كل من نموذجي الانحدار والمخاطرة المتكافئة تتطابق مع نظرية القرار فإن بترسون ونوفيك يؤيدون هذه النماذج بدرجة أكبر، ومع ذلك فقد أشاروا إلى أن كل من هذه النماذج يمثل موقع لقيمة واحدة، وهذا ما لا يمكن تطبيقة دائماً، وستتوضح هذه النقطة أكثر بعد عرض نظرية القرار، أما بالنسبة لنموذج دارلنغتون فإنه لا يتساوى مع نظرية القرار تماماً من وجهة نظر بترسون ونوفيك، مع أن هذا ممكناً، ومع ذلك فإنه يمثل واحدة فقط من نظام القيم العديدة الممكنة.

نظرية القرار كطريقة للاختيار:

يمكن استخدام طريقتين أساسيتين في التحليل في حالة تطبيق نظرية القرار في مشكلات الاختيار، ولتسهيل التفسير سنعرض أولاً صيغ التحليل الموسعة بدلاً من صيغ التحليل الطبيعي (التي تعتمد المنحنى الطبيعي) المستخدمة في أدبيات القياس (Gross & Su, 1975 , Peterson, 1976) والخطوات الأساسية في الصيغة الموسعة للتحليل المستخدمة في نظرية القرار تكون على النحو الآتي:

1- يجب تحديد احتمالية النجاح والفشل لأي مفحوص.

2- يجب أن يؤشر صانع القرار أوزاناً للقيم النسبية (والتي تسمى المستخدمية Utility) لكل نتيجة ممكنة (نجاح أو فشل) لقرار الاختيار.

3- حساب المستخدمية المتوقعة لكل قرار نسبي – القبول أو الرفض –.

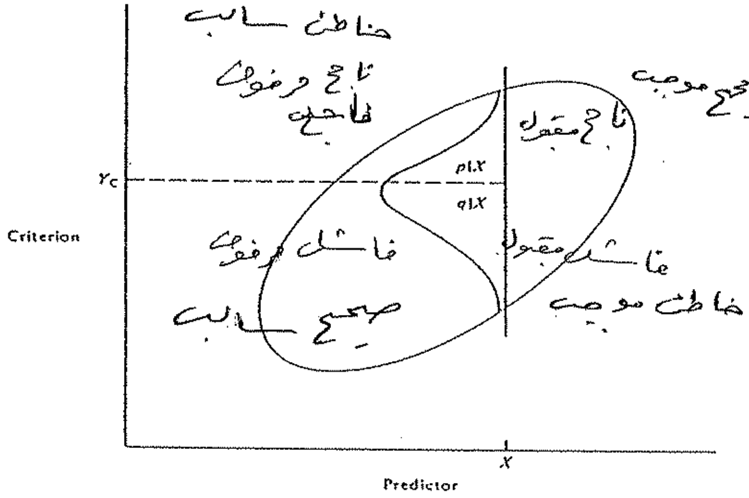
4- مقارنة قيم المستخدمية المتوقعة الأكبر على أنها القرار المناسب.

وسيفسر الجزء الآتي الطرائق والحسابات المستخدمة في الإجراءات المتبعة في الخطوات الأربعة السابقة.

تحديد الاحتماليات والمستخدميات:

دعنا نركز على الاختيار من مجموعة واحدة، افترض أن توزيع ثنائي المتغير لدرجات المتنبئ والمحك للمفحوصين جميعها معروفة من دراسة سابقة (دراسة الصدق)، وأن درجة النجاح على المحك y_2 تحددت بتقسيم المفحوصين لمجموعتين (ناجحة وغير ناجحة) وحيث أن

التوزيع ثنائي المتغير معروف فإن توزيع درجات المحك ممكنة للمتقدم الذي له درجة معينة على المتنبئ، وكنتيجة فإن احتمالية النجاح أو الفشل للمفحوص تصبح معروفة، دع p/x و q/x تمثلان احتماليات النجاح والفشل لدرجة معينة على المتنبئ (انظر شكل 12–5). الكمية px هي المساحة فوق y_2 في التوزيع احادي المتغير فوق x و qx هي المساحة تحت y_2 في التوزيع نفسه.



شكل (12–5): توضيح للاحتمالات المشروطة للنجاح (p/x) وللفشل.(q/x)

تذكر أن هنالك حوادث أربعة لقرار الاختيار صحيح موجب، خاطئ موجب، خاطئ سالب، افترض أنه يمكن ترتيب الاحداث الأربعة من خلال هذه القيم لصانع القرار على سبيل المثال القرارات الصحيح الموجب والصحيح السالب كلاهما احداث ناتجة عن قرارات صائبة، وبالتالي فإنها تكون لها أهمية متساوية كذلك فإن القرارات الخاطئة الموجبة، والخاطئة السالبة كلاهما احداث ناتجة عن قرارات غير صائبة وبالتالي فإنها ذات فائدة أقل من تلك القرارات الصائبة، واعتماداً على وجهة نظر صانع القرار فإن هنالك فائدة أقل للقرارات غير الصائبة أكثر مما هي للقرارات الأخرى. ولتوضيح هذا دعنا نفترض أن الحدث الخاطئ السالب هو الأقل فائدة، وعلى المدى البعيد فإن الاحداث الممكنة يتم ترتيبها حسب قيمتها، افترض الآن أن صانع القرار كان بإمكانة تحديد قيماً عددية للاحداث الأربعة تشير إلى مرغوبيتها النسبة:

$1.00 = \mu_{tp}$

$1.00 = \mu_{tu}$

$0.25 = \mu_{fu}$

$0.75 = \mu_{fp}$

تسمى هذه القيم العددية مستخدميات، ويشير الحرف الأدنى إلى صحيحة موجبة،

صحيحة سالبة، خاطئة سالبة خاطئة موجبة على التوالي، وعند هذه النقطة فلن نتشعب لبحث الطرائق المتبعة للحصول على هذه القيم المشار إليها، ولكن يمكن للقارئ المهتم الرجوع إلى (Novick & Lindsey, 1978).

حساب المستخدمية المتوقعة:

كما هو في نظرية القرار، فإن القرار النسبي عن المفحوص يمكن معرفته بحساب متوسط أو (المتوقع) المستخدمية لكل قرار بديل ممكن، واختيار البديل ذو القيمة المتوقعة الأكبر ولأي قرار يتم اختياره فإن الاحداث الممكنة هي: الصحيح الموجب، والخاطئ الموجب واحتمالية الصحيح الموجب هي $P/_x$ وهي احتمالية النجاح عند درجة المفحوص x، وبالطريقة نفسها فإن احتمالية الخطأ الموجب هي $q/_x$، وهي احتمالية الفشل عند درجة معينة x، وتحسب المستخدمية المتوقعة للقرار المختار بـ:

$$\mu_{fp} (q / x) + \mu_{tp} (p/x) = E (u)_s \quad \text{(2-12)...............}$$

وهو أن المستخدمية المتوقعة لقرار مختار تحسب بضرب مستخدمية كل حدث باحتماليته، وجمع نتائج الاحداث الممكنة جميعها لذلك القرار، والمستخدمية المتوقعة لقرار الرفض هي:

$$\mu_{tn} (q/x) + \mu_{fn} (p/x) = E (u)_r \quad \text{(3-12)..................}$$

وقد يندهش القارئ لأن الاحتمالية المحسوبة من كلا المعادلتين هي نفسها، افترض أن p/x تمثل احتمالية النجاح لدرجة معنية x، ولاختيار قرار فإن المفحوص الناجح عبارة عن حدث صحيح موجب، وعلى هذا فإن احتمالية الصح الموجبة تكون $P/_x$، وμ_{TP} تضرب بـ $p/_x$.

والقرار بالرفض للمفحوص الناجح عبارة عن حدث خاطيء سالب، وعلى هذا فإن احتمالية، الخطأ السالب هي $p/_x$ و μ_{Fn} تضرب بـ $p/_x$ عند حساب $E(u)_r$.

ولتوضيح هذا، افترض كما هو سابقاً أن $\mu_{tp} = \mu_{tn} = 1$، $\mu_{fn} = 0.25$، $\mu_{fp} = 0.75$، وافترض أنه لمفحوص ما، كانت $P/_x = 0.4$ و $q/x = 0.6$ لذلك فإن:

$$E(u)_s = 1 \times 0.4 + 0.75 \times 0.6 = 0.85$$

$$E(u)_r = 0.25 \times 0.4 + 1 \times 0.6 = 0.70$$

وحيث أن المستخدمية المتوقعة لاختيار القرار تتجاوز المستخدمية المتوقعة لرفضه فإنه سيتم اختيار المفحوص، وهذا صحيح، ومع أن احتمالية فشل المفحوص أكبر من احتمالية

نجاحه، والقرار المختار مفضل لأن مستخدمية الحدث الخاطئ الموجب ثلاثة اضعاف مستخدمية الحدث الخاطئ السالب.

تطبيق نظرية القرار في الاختيار بين مجموعتين:

عندما تستخدم أكثر من مجموعة في عملية الاختيار فإن احتماليات نجاح المفحوصين على المتنبئ نفسه تكون مختلفة اعتماداً على المجموعة التي ينتمون إليها، وعلاوة على ذلك فإن صانع القرار يأمل بتأشير قيم مستخدمية مختلفة للنتائج المكنة المتنوعة للاختيار من مجموعات المتقدمين المختلفة، وسنوضح الآن كيف يمكن تطبيق صيغ التحليل الشاملة لاختيار القرار باستخدام مجموعتين مختلفتين من المتقدمين. عندما يكون لدينا مجموعة واحدة، فإن هنالك أربعة أحداث ممكنة، ومع مجموعتين يوجد ثمانية احتمالات ممكنة: صحيحة موجبة للمجموعة الرئيسية، صحيحة موجبة للمجموعة الثانوية... وهكذا، وكنتيجة توجد مستخدمية لثمانية احداث ممكنة، مثلاً افترض أن المستخدمية للمجموعة الرئيسية هي $\mu_{tp1}=1$ و $\mu_{rn}=0.25$ و $\mu_{fn1}=0.5$ ، وافترض أيضاً أن المستخدمية للمجموعة الثانوية هي $M_{tp2}=1.25$ و $\mu_{tn2}=0.25$ و $\mu_{fn2}=0.5$ وتتضمن هذه المستخدميات قبول مفحوص المجموعة الثانوية الذي سينجح والذي تكون له قيمة أكبر من تلك لقبول مفحوص في المجموعة الرئيسية (وتتضمن أيضاً أن قبول المفحوص الناجح أكثر قيمة من رفض المفحوص الذي سيفشل في المجموعة الثانوية).

افترض وجود مفحوصين اثنتين على المتنبئ نفسه أحدهما في المجموعة الرئيسية والأخر في المجموعة الثانوية، وافترض أن احتمالية النجاح لمفحوص المجموعة الرئيسية هي $p_1/x = 0.55$، واحتمالية الفشل $q_1/x = 0.45$ ولمفحوص المجموعة الثانوية فإن الاحتماليات المماثلة هي $p_2/x = 0.52$ و $q_2/x = 0.48$ وعلى هذا فإن احتمالية النجاح تكون أكبر للمفحوص في المجموعة الرئيسية وتكون قيمة المستخدمية له:

$$\mu_{fp1}(q_1/x)+\mu_{tp1}(p_1/x)=E(u)_{s1}$$

$$(0.45)0.25+(0.55)1=$$

$$0.6625=$$

$$U_{tn1}(q_1/x)+U_{fn1}(p_1/x)=E(u)_{r1}$$

$$(0.45)1+(0.55)0.5=$$

$$0.725=$$

ومما سبق فإن مفحوص المجموعة الرئيسية سيرفض لأن قرار الرفض له قيمة مستخدمية أكبر.

وللمفحوص في المجموعة الثانوية تكون قيمة المستخدمية:

$$U_{fp2}\ (q_2/_x) + U_{tp2}\ (p_2/_x) = E\ (u)_{s2}$$

$$(0.48)0.25 + (0.52)\ 1.25 =$$

$$.0.77 =$$

و $$Ut_{n2}\ (q_2/_x) + U_{fn2}\ (p_2/_x) = E\ (u)_{r2}$$

$$(0.48)\ 1 + (0.52)\ 0.5 =$$

$$.0.74 =$$

وعلى هذا، فإن المتقدم من المجموعة الثانوية سيتم قبوله، وهذا قرار صائب مع أن المفحوصين لهما الدرجة نفسها على المتنبيء، وتكون فرصة نجاح المفحوص من المجموعة الثانوية أقل من فرصة نجاح مفحوص المجموعة الرئيسية وهذا يعود إلى أن مستخدمية القبول الصحيح للمجموعة الثانوية تساوي 1.25 ضعف ذلك للمجموعة الرئيسية (وذلك للمستخدمية الصائبة للمجموعة الرئيسية).

مفاهيم إضافية:

كانت المناقشة السابقة لنظرية القرار مدخلاً لاستخدامها في مشكلات الاختيار ولفهم كامل وأفضل لمستخدميها الجدد فإنه يتطلب منهم معرفة مفاهيم إضافية عدة، والتي ستقدم في هذا الفصل.

وكما أشرنا سابقاً فإن ما يطلق عليه صيغ التحليل الطبيعية أكثر شيوعاً من صيغ التحليل الموسع في أدب القياس، وفي طريقة صيغة التحليل الموسع يتم حساب المستخدمية والرفض المتوقع لقرارات الاختيار للمتقدم، ويكون القرار لأكبر قيمة ممكنة للمستخدمية المتوقعة للمتقدم، ويتم إجراء حسابات منفصلة لكل مفحص، وتركز صيغ التحليل الطبيعية على المستخدمية المتوقعة ولكنها مصممة لتحديد درجة قطع المتنبئ التي تعطي أكبر قيمة ممكنة للمستخدمية المتوقعة، وعندما تتحدد درجة القطع يمكن اتخاذ قرار بخصوص المتقدم وذلك بمقارنة درجته على المتنبئ بدرجة القطع وهنا لا يكون من الضروري اجراء حسابات للمستخدمية المتوقعة الخاصة بكل متقدم.

ولإجراء تحليل الصيغة الطبيعية دعنا نركز ثانية على الاختيار من مجموعة واحدة، وسنفترض ثانية توزيع ثنائي المتغير لدرجات المتنبئ والمحك والمعروفة من أبحاث سابقة، وكذلك حددت درجة النجاح على المحك، وللأهداف المتصورة افترض أن معرفتنا عن التوزيع تتمثل في خط بياني مثل ذلك في شكل (12–1ب) (وفي البحوث النظرية مثل تلك الممثلة بابحاث غروس وسو (Gross & Su, 1975) يفترض أن يكون التوزيع الثنائي الطبيعي، ولا يركز التحليل على مجموعة بيانات معينة ولكن على التوزيع المفترض)، وعند تحديد درجة القطع على المتنبئ فإن التمثيل البياني يصبح كما في شكل (12–1د)، ويمكن تحديد احتمالية الاحداث الأربعة، الصحيحة الموجبة، والخاطئة الموجبة، والصحيحة السالبة، والخاطئة السالبة، لاحظ أن هذه الاحتمالات تحسب لمجموعة المفحوصين الكلية مقابل حساب هذه الاحتمالات لكل مفحوص في صيغ التحليل الموسعة، ويرمز لاحتمالات الحوادث الأربعة بـ:

P_{fn} , P_{tn} , P_{fp} ,P_{ip}, وبالتالي تكون المستخدمية المتوقعة لدرجة قطع معينة والتي تعتمد على الاحتمالية هي:

$$\mu_{fn}P_{fn} + \mu_{tn}P_{tn} + \mu_{fp}P_{fp} + \mu_{tp}P_{tp} = E\,(U) \qquad \text{(4-12)................}$$

لاحظ أن المستخدمية المتوقعة هي الجمع الموزون لمستخدمية الاحداث الأربعة، والأوزان هي احتمالات هذه الحوادث لاحظ أيضاً أن الفرق بين المعادلة 12 – 4 والمعادلة 12 – 2 أو 12 – 3 أن الأولى كانت للمفحوصين جميعهم في حين أن الأخيرة هي لمتقدم (مفحوص) معين.

وكما وضحنا في اشتقاق المعادلة (12–4) فإن المستخدمية المتوقعة تحسب لدرجة قطع معينة، وهدف صيغ التحليل الطبيعي هو تحديد درجة القطع التي تعطي أقصى قيمة للمستخدمية المتوقعة، ولأهداف مفاهيمية قد تفكر بأن هذه العملية هي نفسها في حساب E (u) لأي درجة قطع ممكنة، ومن ثم اختيار درجة القطع التي تعطي أكبر قيمة لـ(u) E ، وعملياً فمن الضروري وجود صيغة لتحديد درجة القطع والتي تعطي أكبر قيمة لـ E (u). إن تحليل الصيغ الطبيعي يمكن استخدامه مع درجات قطع منفصلة لكلا المجموعتين الرئيسية والثانوية والتي تعطي أقصى قيمة للمستخدمية المتوقعة، وتذكر أن هنالك ثمانية احداث، وتؤدي صيغ التحليل الطبيعي إلى أقصى مستخدمية متوقعة عبر الاحداث الثمانية جميعها، والتعبير الذي يجب أن يُعطي أكبر قيمة هو:

$$(\mu_{tni}\, Ptni + \mu\, f_{ni}\, P_{fni} + \mu\, f_{pι}\, P_{fpι} + U_{tpi}\, P_{cpi})\, P_i \sum^{2} = E(u) \qquad \text{(5-12)......}$$

والمصطلح الوحيد الجديد هو Pi(i = 2.1) ويمثل نسب الطلبة في المجموعات الرئيسية

والثانوية على التوالي، وتكبير هذا التعبير يكافئ تعديل درجتي القطع واحدة للمجموعة الرئيسية وأخرى للمجموعة الثانوية، ويستمر هذا حتى نصل إلى أقصى قيمة للمستخدمية المتوقعة، وللتفاصيل يمكن الرجوع إلى (Cross & Su, 1975) وذلك في حالة افتراض التوزيع الطبيعي لمتغير ثنائي (توزيع المحك والمتنبئ). وبيّن **بترسون** (Peterson, 1976b) أن الطريقة التي طورها كل من غروس وسو يمكن استخدامها كمبدأ مع توزيعات أخرى.

محدد أخر للتغطية في هذا الفصل هي مناقشة ما يسمى دالة عينة المستخدمية ففي دالة المستخدمية تكون مستخدمية أي من الاحداث الحقيقية الموجبة مثلاً مستقلة عن درجات المحك الفعلية التي حصل عليها المفحوص أو بكلمات أخرى، فعند نجاح مفحوصين اثنين فإن مستخدمية قرار الاختيار تكون هي نفسها بغض النظر عن أي فرق بين درجات المحك للمفحوصين، ويبدو أن هذا منطقياً في العديد من المواقف، أن مستخدمية تاجر أدت جهوده إلى ربح 2 مليون دولار بالتأكيد أكبر من مستخدمية تاجر أخر أدت جهوده إلى ربح 50.000 دولار (وللمهتم بحساب أثر استدلال عتبة المستخدمية انظر Schmidt & Hunter, 1981)

لهذا فإن دوال المستخدمية غير دالة لعتبة المستخدمية والتي تؤخذ بعين الاعتبار في مشكلات الاختيار. وأخيراً كما تم ملاحظته مسبقاً فإن قضية اعطاء قيم عددية للمستخدمية لم تتم عنونتها، وكل من مجموعة المستخدمية ودوال المستخدمية تم بحثها في (Noviek & Lindly, 1978) والذي قدم عرضاً لعدد كبير من المراجع في هذا الموضوع.

الخلاصة:

مع أنه قد يتوافر للاختبار مؤشر تجريبي لصدقه (في صيغة ارتباط بين درجات الاختبار ودرجات المحك) فإن هنالك اعتبارات إضافية لتقييم فائدة درجات الاختبار لقرارات الاختيار، أحد هذه الاعتبارات تستخدم محدِد في موقف الاختيار نفسه، بشكل خاص فلأي نسبة قاعدية لمجتمع المفحوصين تتفاعل عوامل صدق درجة الاختبار ونسبة الاختيار لتؤثر في نسبة النجاح التي نحصل عليها، ويمكن تبسيط هذه باستخدام القيم المستمدة من جداول تايلور – روشل، ومن الواضح أنه إذا كان هنالك مجموعتين مختلفتين من المفحوصين لهما نسب نجاح قاعدية مختلفة فإن استخدام نسبة الاختيار الشائعة واختيار معاملات صدق متماثلة لكلا المجموعتين ينتج عنها اتوماتيكياً نسب نجاح مختلفة.

قضية أخرى مهمة مرتبطة باستخدام الاختبار لأهداف الاختيار هو، إذا كان قرار الاختيار عادلاً لمجموعات مفحوصين مختلفة، ولتناول هذه القضية فقد طوّر السيكومتريين تعريفات مختلفة لتحيز الاختبار في الستينات، وعرضت تعريفات لكل من كليري وثورندايك ودارلنغتون، ونقطة مهمة متعلقة بهذه التعريفات هو أن الاختبار غير متحيز بهذه التعريفات واختيار النماذج المشتقة منها يجب تتبع تطويرها مباشرة.

ومع ذلك فإن نماذج عادلة متنوعة برزت من هذه التعريفات لتحيز الاختبار والتي لها أساس في نظرية الاختيار التي تمت مراجعتها في هذا الفصل تتضمن نموذج الانحدار، ونموذج المخاطرة المتساوية، ونموذج دارلنغتون، ونموذج الاحتمالية المتساوية، ونموذج النسبة الثابتة، ونموذج الاحتمالية المشروطة، ونموذج الاختيار النظري ويشترط الأخير من مستخدم الاختبار أن تكون له القدرة على تأشير قيمة موزونة لكل نتيجة محتملة لقرار الاختيار: الصحيح الإيجابي، الخطأ الإيجابي، الصحيح السلبي، الخطأ السلبي، وبمعرفة احتمالية النجاح لشهادة التأهيل لدرجة معينة على الاختبار المتنبئ، فإن المستخدمية المتوقعة لقرار الاختيار لقرار الرفض يجب حسابها فيما بعد. والقرار البديل ذا المستخدمية الصح الإيجابية الأكبر هو الاختيار لدرجة المتنبئ المعينة نفسها، ومجموعة منفصلة من أوزان المستخدمية يمكن تأسيسها لكل مجموعة مفحوصين في تمييز القيمة التفاضلية للمؤسسة أو المجتمع هو ظهور كل من الصحيح الايجابي، والصحيح السلبي، والخطأ الايجابي والخطأ السلبي لمجموعات المفحوصين المختلفة، ووضحت هذه التطبيقات من خلال استخدام صيغ التحليل الشاملة.

ونقطة أخيرة حول هذه النماذج للاختيار العادل التي اقترحت، هو أنها جميعها قد تعكس قيم اجتماعية مختلفة، ويوصى بنموذج القرار النظري عبر الأخرى لأنها تسمح لمستخدم الاختبار ليأخذ بعينة الاعتبار الاحداث المحتملة جميعها التي تظهر في عملية الاختيار لأنها تتطلب جمل عامة واضحة للقيم المؤشرة للحوادث لكل مجموعة مختلفة من المفحوصين، ولأنه يزود قانون قرار يستخدم تجمع من الاحتمالات لهذه الاحداث وقيمها النسبية لصانع القرار، كذلك تم تحديد موضوعات اضافية في نموذج الاختيار النظري والتي قد تهم بعض القارئين.

التمارين:

1/ اختير عشوائياً (10) من الذكور و (10) من الإناث من مجموعة كبيرة من الطلبة للاشتراك في برنامج تدريبي في العلوم في المدرسة الثانوية، وفي بداية البرنامج الذي استمر لسنة طبق على الطلبة اختبار تحصيلي في العلوم ودونت درجات الطلبة في الاختبار ومتوسطهم في البرنامج والاحصائيات الوصفية لهؤلاء الطلبة.

الذكور			الاناث		
المفحوص	درجة الاختبار	المتوسط	المفحوص	درجة الاختبار	المتوسط
1	92	3.00	11	91	3.70
2	84	2.80	12	78	1.90
3	96	3.60	13	88	3.20
4	58	2.00	14	69	1.20
5	75	2.50	15	68	2.50
6	94	3.40	16	66	3.00
7	78	2.70	17	82	3.40
8	55	1.20	18	62	1.90
9	72	1.20	19	65	2.40
10	76	1.60	20	68	2.00

المتوسط	77	2.40	74	2.50
الانحراف المعياري	13	0.83	10	0.79
الارتباط	0.87		0.78	

أ) اقترح مطبقوا البرنامج استخدام درجة قطع (70) في اختبار العلوم التحصيلي لاختيار الطلبة في البرنامج مستقبلاً، وحدَّد النجاح في البرنامج على أنه متوسط نقاط 2.00 باستخدام البيانات الملائمة. حدد ما إذا كانت طريقة الاختيار هذه عادلة وفقاً للنماذج الآتية:

1- الاحتمالية المتساوية.

2- النسبة الثابتة

3- الاحتمالية الشرطية

4- نموذج الانحدار

ب) افترض أن برنامج السنة الثانية تم تكييفه على30% من المفحوصين، ما درجة القطع المناسبة لكل من الذكور والإناث لاختيار طريقة عادلة وفقاً لنموذج الاحتمالية المشروطة.

ج) ما درجة قطع المتنبئ التي يجب انشاءها لكل من الذكور والإناث فيما لو استخدم نموذج الانحدار ودون أي حد لعدد الطلبة الذين سيقبلوا في البرنامج.

د) على وفق تعريف ثورندايك للاختبار غير المتحيز هل يعد اختبار العلوم التحصيلي اختبار عادل؟

2/ أدت دراسة صدق لمتنبئ مفرد إلى الاحصاءات الآتية لكلا المجموعتين الرئيسية والفرعية.

الاحصائي	المجموعة	
	الرئيسية	الفرعية
القاطع	40.0	45.0
الميل	1.5	1.0
الخطأ المعياري للتقدير	4.0	5.0

فإن كانت درجة النجاح على المحك 60، ويرغب المعهد في الموافقة على المتقدمين الذين حصلوا على فرصة 75% للنجاح على الأقل، ما درجات القطع التي يجب وضعها للمجموعات الرئيسية والفرعية.

3/ افترض أن المستخدمية للمتقدمين في المجموعات الرئيسية والفرعية على النحو الآتي:

المجموعة		
الحدث	الرئيسية	الثانوية
صحيح إيجابي	1.00	1.50
صحيح سلبي	1.00	1.00
خطأ إيجابي	0.50	0.75
خطأ سلبي	0.50	0.25

- واحتمالات النجاح والفشل للمتقدمين في كلا المجموعتين على درجة متنبئ معين كانت كما يأتي:

المجموعة		
النتائج	الرئيسية	الثانوية
النجاح	0.60	0.5
الفشل	0.40	0.5

- فما القرارات التي يجب عملها للمتقدمين في كلا المجموعتين على درجة هذا المتنبئ؟

الفصل الثالث عشر

13

التحليل العاملي

الفصل الثالث عشر

التحليل العاملي

عند تطبيق اختبارات عدة على مجموعة المفحوصين نفسها، فإن احد مظاهر التصديق قد يستخدم تحديد ما إذا كان هنالك واحداً أو أكثر من تجمع للاختبارات يؤدي المفحوصين من خلالها أداء متشابه نسبياً. ويتحدد عدد وتركيب التجمعات بالارتباطات بين أزواج الاختبارات جميعها. على سبيل المثال إذا كانت بطارية اختبارات الذاكرة تتألف من عدة اختبارات تحوي مثيرات ذات معنى في الفقرات واختبارات أخرى لا تحويها، فقد يكون من المثير رؤية ما إذا كان نمط الارتباطات عبر الاختبارات تدعم فرضية أن هنالك تجمعين مختلفين لاختبارات الذاكرة. أو تخيل دراسة عدة سمات – عدة طرق التي يفترض أنها تقيس ثلاث سمات، هنا قد يكون من المفيد رؤية ما إذا كانت الارتباطات تدعم فرضية ثلاث سمات مختلفة. إحدى الطرق المناسبة للتحقق من هذه الأسئلة هو التحليل العاملي .

يهدف هنا الفصل إلى تزويدنا بمقدمة شاملة لعملية التحليل العاملي، كما تطبق في دراسة الصدق الاكتشافي لبطارية اختبارات أو بعض المجموعات الأخرى من القياسات. أولاً سنقدم مثال مع بيانات افتراضية وتفسير المفاهيم الأساسية للتحليل العاملي ضمن السياق والعرض المقدم يساعد القارئ في الإجابة عن الأسئلة الآتية :

1/ كيف يشير نمط الارتباطات عبر المتغيرات في مصفوفة الارتباطات إلى عدد العوامل المؤثرة في الأداء ؟

2/ ماذا نعني بتشبعات العوامل؟

3/ كيف تستخدم تشبعات العوامل في تحديد عدد العوامل (أو تجمعات الارتباطات) في مصفوفة الارتباط الملاحظة؟

4/ ما هي الدورات؟

5/ كيف يعرف الباحث أي مجموعة من تشبعات العوامل (الناتجة عن الدورات المختلفة) لتفسيرها ؟

6/ كيف تعزى مفاهيم الطائفية والتفرد إلى الثبات؟

وسنعرض في النصف الأخير من هذا الفصل مثال كامل للتحليل العاملي لدرجات اختبارات فرعية لبطارية اختبارات ذكاء لتوضيح تفسير نتائج سياق عملية التصديق الواقعية. وينتهي الفصل بمقدمة مختصرة للتحليل العاملي التوكيدي .

مثال مع بيانات افتراضية:

افترض توافر درجات مجتمع مفحوصين على الاختبارات الفرعية لاختبار وكسلر للمراهقين، والمؤلف من احد عشر اختباراً فرعياً.

ستة منها لفظية وخمسة أدائية. وعلى أساس تصنيف مجموعات الاختبارات الفرعية فقد يفترض احدنا إن الارتباطات عبرها تقترح تجمعين للاختبارات الفرعية.ويمكن إثبات هذا بالارتباطات الافتراضية عبر الاختبارات الفرعية المبينة في جدول (13-1). وهنا كل زوج من الاختبارات اللفظية تظهر ارتباط داخلي أساسي، وكذلك كل زوج من الاختبارات الأدائية. ومع ذلك فان الارتباط بين أي زوج من الاختبارات الأدائية – اللفظية فقط 15 و0، وفي لغة التحليل العاملي فإن الاختبارات الفرعية تقيس عاملين، ففي مجموعة الاختبارات اللفظية تقيس عامل واحد، وفي المجموعة الأدائية تقيس عامل ثاني، بالمقابل اذا كانت الارتباطات بين الاختبارات جميعا = 0.6 مثلاً، فان هذه الارتباطات تشير إلى تجمع واحد للاختبارات، أي إن الاختبارات تقيس عامل واحد.

العامل وتشبعات العوامل :

العامل هو متغير غير ملاحظ أو كامن مثل الدرجة الحقيقية غير الملاحظة، وكما يمكن حساب معامل الارتباط بين الدرجة الملاحظة والدرجة الحقيقية فمن الممكن حساب معامل الارتباط بين الاختبارات والعوامل . ويبين جدول (13-2) الارتباطات بين اختبارات وكسلر الفرعية والعوامل المشار إليها في جدول (13-1) .

جدول (13-1) : حسابات افتراضية عبر اختبارات وكسار الفرعية :

المقياس	الاختبار الفرعي	1	2	3	4	5	6	7	8	9	10	11
اللفظي	1. المعلومات	-										
	2. الفهم	0.65	-									
	3. الحساب	065	0.65	-								
	4. المتشابهات	0.65	0.65	0.65	-							
	5. سلاسل الأعداد	0.65	0.65	0.65	0.65	-						
	6. الألفاظ	0.65	0.65	0.65	0.65	0.65	-					
الأدائي	7. رموز الأعداد	0.15	0.15	0.15	0.15	0.15	0.15	-				
	8. إكمال الصورة	0.15	0.15	0.15	0.15	0.15	0.15	0.50	-			
	9. تصميم المكعبات	0.15	0.15	0.15	0.15	0.15	0.15	0.50	0.50	-		
	10. ترتيب الصور	0.15	0.15	0.15	0.15	0.15	0.15	0.50	0.50	0.50	-	
	11. تجميع الأشياء	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.50	0.50	0.50	-

جدول (13-2): تشبعات العوامل الافتراضية لاختبارات وكسلر الفرعية:

المقياس	الاختبار الفرعي	العامل	
		الأول	الثاني
اللفظي	1. المعلومات	0.8	0.1
	2. الفهم	0.8	0.1
	3. الحساب	0.8	0.1
	4. المتشابهات	0.8	0.1
	5. سلاسل الأعداد	0.8	0.1
	6. الألفاظ اللغوية	0.8	0.1
الأدائي	7. رموز الأعداد	0.1	0.7
	8. إكمال الصور	0.1	0.7
	9. تصميم المكعبات	0.1	0.7
	10. ترتيب الصور	0.1	0.7
	11. تجميع الأشياء	0.1	0.7

ويطلق على هذه الارتباطات اسم تشبعات العوامل، فكل ارتباط لفظي فرعي يرتبط مع العامل الأول ب 0.8 ومع العامل الثاني ب 0.1 فقط. مع الأخذ بعين الاعتبار إن الاختبارات التي ترتبط مع العامل الأول هي الاختبارات اللفظية (أي العامل اللفظي) والتي ترتبط مع العامل الثاني هي الاختبارات الأدائية (أي العامل الأدائي) والارتباط بين زوج من الاختبارات له علاقة مهمة بتشبعات الاختبارين على العوامل، وعندما يكون هنالك عاملين اثنين فان العلاقة تكون

$$a_{j2}\, a_{i2} + a_{j1}\, a_{i1} = \rho_{ij} \qquad (1\text{-}13) \ldots\ldots\ldots\ldots$$

ويمثل: ρ_{ij} الارتباط بين الاختبارين i و j، و a_{j1} تشبعات الاختبارين i و j بالعامل الأول و a_{j2} , a_{j2} تشبعات الاختبارين i و j بالعامل الثاني. كتوضيح، افترض إن الارتباط بين اختباري المعلومات والرموز العددية $\rho_{17} = 0.10$، وتشبعات العوامل $a_{11} = 0.8$ ، $a_{12} = 0.1$ ، $a_{71} = 0.1$ ، $a_{i2} = 0.7$ بالتعويض في المعادلة (13-1) يؤدي الى:

$$0.15 = (0.7)\, 0.1 + (0.1)\, 0.8 = \rho_{17}$$

والمعادلة العامة التي تربط الارتباطات الداخلية بتشبعات العوامل هي:

$$a_{jk}\; a_{ik} \sum_{l=i}^{m} = \rho_{ij} \qquad (2\text{-}13) \ldots\ldots\ldots\ldots$$

حيث ترمز m إلى عدد العوامل.

الدوران:

تعد المعادلة (13-1) وتعميمها بالمعادلة (13-1) معادلات مهمة في فهم دوران العوامل. بمتابعة مثال اختبارات وكسلر الأحد عشر الفرعية افترض مجموعة تشبعات عوامل جديدة وكما هو مبين في جدول (13-3) مع إن هذه المجموعة مختلفة تماماً عن المجموعة المبينة في جدول (13-2)، وكلا المجموعتين تحقق المعادلة (13-1)، فمثلاً ترميز تشبعات العوامل في جدول (13-3) بـ: $a^*_{j2}\;\; a^*_{j1}\;\;\; a^*_{i2}\;\;\; a^*_{i1}$ فإن الارتباط بين اختباري المعلومات والرموز العددية هو :

جدول (13-3): تشبعات العوامل البديلة لاختبارات وكسلر الفرعية

المقياس	الاختبار الفرعي	العامل	
		الأول	الثاني
اللفظي	1. المعلومات	0.6363	0.4949
	2. الفهم	0.6363	0.4949
	3. الحساب	0.6363	0.4949
	4. المتشابهات	0.6363	0.4949
	5. سلاسل الاعداد	0.6363	0.4949
	6. الألفاظ اللغوية	0.6363	0.4949
الأدائي	7. رموز الأعداد	0.5656	0.4242-
	8. إكمال الصورة	0.5656	0.4242-
	9. تصميم المكعبات	0.5656	0.4242-
	10. ترتيب الصور	0.5656	0.4242-
	11. تجميع الأشياء	0.5656	0.4242-

$$a^*_{72}\ \ a^*_{71}+\ \ \ a^*_{12}\ \ \ a^*_{11} = \rho_{17}$$

$$0.15 = (0.2424-\ )\ 0.4949 + (0.5656)\ 0.6363 =$$

وبما أن تشبعات العوامل في كلا الجدولين (13-2) و (13-3) تحقق المعادلة (13-1)، فمن غير الممكن الاختيار بين المجموعتين على أساس ارتباطات الاختبارات الفرعية. وفي الحقيقة هنالك عدد لا نهائي من مجموعات تشبعات العوامل تحقق جميعها المعادلة (13-1)، وأي واحدة من هذه المجموعات يمكن الحصول عليها من مجموعة ثانية خلال عملية تسمى الدوران وجاء هذا المصطلح من التفسير الهندسي للتحليل العاملي (وسنقدم توضيحاً هندسياً للدوران لاحقاً في هذا الفصل).

وللمثال الافتراضي سنقدم فوراً توضيح هندسي الذي سيكون مربكاً نوعاً ما، لأن العوامل الفرعية الستة الأولى، وجميعها له تشبع العامل الثاني نفسه، وجميعها له تشبع العامل الثاني نفسه، لذا فإنه في أي عرض هندسي من غير الممكن التمييز عبر العوامل الستة الأولى أو عبر الخمسة الأخيرة. وترتبط الدورانات المختلفة بتحويلات رياضية، على سبيل المثال ترتبط تشبعات العامل في جدول (13–2) وجدول (13–3) كما يأتي:

$$0.707a_{i2} + 0.707\ a_{i1} = a^*_{i1}$$ .... (13-3 أ)

$$0.707a_{ik} + 0.707\ a_{i1} = a^*_{i2}$$ .... (13-3 ب)

للتوضيح، افترض إن اختبار الرموز العددية الفرعي له تشبع عامل $a_{71} = 0.1$، $a_{72} = 0.7$. وبالتعويض في المعاملات 13–3أ و 13–3 ب نحصل على :

$$0.5656 = (0.7)\ 0.707 + (0.1)\ 0.707 \quad = \quad a_{71}$$

و

$$0.4242 = (0.7)\ 0.707 - (0.1)\ 0.707 \quad = \quad a^*_{72}$$

إن عملية الحصول على الأعداد التي تحول مجموعة من التشبعات إلى الأخرى ليست مهمة في الحقيقة. والنقطة المفتاحية هي انه عند الحصول على مجموعتين مختلفتين من تشبعات العوامل بعملية الدوران تتكون كل دورة من التشبعات على العدد نفسه من العوامل التي تعزى إلى الارتباطات بالتساوي، وحجم تشبعات العوامل هو الذي يتغير فقط. كذلك على القارئ ان يلاحظ ان هنالك مجموعة مختلفة من العوامل التي تعزى إلى كل مجموعة من التشبعات . ومع إن الدوران يغير العوامل أو المتغيرات غير الملاحظة، مما يشير الى إن درجات الاختبار يبدو أنها مرتبطة بذلك .

وحقيقة إن المجموعات المختلفة من تشبعات العوامل تحقق المعادلة (13–2) لمصفوفة الارتباطات نفسها لا تكون مشكلة فيما إذا كانت مجموعات التشبعات لها تفسيرات متشابهة، ولسوء الحظ فإن هذا ليس صحيحاً على وجه العموم، فمثلاً مجموعتي تشبعات العوامل في الجدولين (13–2) و (13–3) تقترح تفسيرات مختلفة للعاملين المقاسين بمقياس وكسلر.تقترح تشبعات جدول (13–2) عامل لفظي وآخر أدائي. وبفحص التشبعات في جدول (13–3) نرى إن الاختبارات جميعها لها تشبع موجب وأساسي على العامل الأول، ويقترح هذا النمط إن العامل الأول عامل عام يقاس باختبارات وكسلر الفرعية جميعها

وبالعودة إلى العامل الثاني سنرى أن الاختبارات اللفظية لها تشبعات موجبة، في حين إن الاختبارات الأدائية لها تشبعات سالبة، وحيث إن كلا الاختبارات الفرعية اللفظية والأدائية ليست مرتبطة سلبياً فان نمط التشبعات على العامل الثاني يقترح إن هذا العامل يهتم بالاختلافات في القدرات المقاسة بالاختبارات اللفظية والأدائية.

وتقترح مجموعتي تشبعات العوامل تفسيرات مختلفة وذلك لوجود مجموعة من العوامل المناظرة لكل مجموعة من التشبعات، وللعاملين الأول والثاني المقابلة للتشبعات في جدول (13–2) نرمز لها ب f_1 و f_2 والعوامل $1^/$ و$2^/$ في حين إن التشبعات في جدول (13–3) يرمز لها ب $f1^*$ و $f2^*$ ويمكن تبيان إن العلاقة بين مجموعتي العوامل هي :

$$(f_2 + f_1)\ 0.707 = f_2\ 0.707 + f_1\ 0.707 = f^*_1$$

و

$$(f_1 - f_2)\ 0.707 = f_2\ 0.707 + f_1\ 0.707 = f^*_2$$

لذا فإن f^*_1 العامل المناظر للعمود الأول للتشبعات في جدول (13–3) هو بالأساس مجموع العاملين f_1و f_2 المناظرة للتشبعات في جدول (13–2) (ويخدم العدد 0.707 ببساطة ليجعل f^*_1 لها تباين =1) وكنتيجة فإنها تجعل التشبعات في العمود الأول تقترح عامل عام. وبعدها فإن كان f_1 يمثل العامل اللفظي و f_2 يمثل العامل الأدائي $f^*_{19} = (f2+f1).707$ يمثل العامل العام الذي تقاس بكل اختبار فرعي في WAIS. بصورة مشابهة f_2^* هو العامل المناظر للتشبعات في العمود 2 في جدول (13–3) وهو بالأساس الفرق بين العاملين المناظرين للتشبعات في جدول (13–2) وثانية تجعل من البديهي إن التشبعات في العمود 2 في جدول (13-3) تقترح أن العامل 2 يتعلق بالفروقات في القدرة المقاسة بالاختبارات الفرعية اللفظية والأدائية.

واذا كان هنالك عدد لا نهائي من مجموعات تشبعات العوامل التي تحقق المعادلة (13–2) لمصفوفة ارتباطات معينة، فكيف للباحث إن يختار مجموعة واحدة من تشبعات العوامل لتفسيرها؟ كمبدأ المجموعة الأكثر ملائمة هي التي تقابل بصورة أفضل محك البناء البسيط الذي طوره ثيرستون عام 1947، ويشير هذا المحك إلى إن كل اختبار له تشبعات كبيرة لأقل عدد ممكن من العوامل، ومنخفضة أو صفر لتشبعات العوامل المتبقيــة. وقد قدم موليك

(mulaik, 1972) وصف أكثر اكتمالاً للبناء البسيط، لذا فإن محك البناء البسيط يفضل التشبعات في جدول (13–2) على تلك في جدول (13–3).

وعند إجراء التحليل العاملي فإن المجموعة الأولى من تشبعات العامل يتم الحصول عليها باستخدام طريقة تسمح بالحساب المناسب لهذه التشبعات. . وهذه التشبعات تسمى الابتدائية أو غير المدورة ,وعادة لا يحاول الباحث تفسير التشبعات غير المدورة والطرائق جميعها ينتج عنها مجموعة من معاملات التحويل مثل تلك التي في المعادلة (13–3) وتستخدم هذه المعادلات لتحويل تشبعات العامل الابتدائي وبالتالي تؤدي لتقريب البناء البسيط، وبعدها يتم تفسير التشبعات المدورة. وهنالك نوعان من الدوران: غير المباشر (oblique) والمستقل (orthogonal) وكما تشير الأسماء فإن الدوران المستقل يؤدي إلى عوامل غير مرتبطة، في حين إن الدوران غير المباشر يؤدي إلى عوامل مرتبطة والتي ستناقش في الجزء التالي. وتجد المناقشة المفصلة لطرق الدوران موجودة في هارمون (harmon 1967) ومولياك (muliak, 1972) والطريقة الأكثر شيوعاً للدوران المستقل هي طريقة تعظيم التباين (varimax (Kaiser , 1958، وطور جينريش وسامبسون (jennrich & Sampson, 1966) طريقة التجزئة الرباعية وهي الأكثر شيوعاً للدوران غير المباشر

العوامل المرتبطة

العوامل عبارة عن متغيرات غير ملاحظة لكنها مثل أية متغيرات أخرى قد تكون مرتبطة، وهو إن بعض الدوران لتشبعات العامل ينتج عنها عوامل مرتبطة، والبعض الآخر ينتج عنه عوامل غير مرتبطة. وتسمى مجموعة تشبعات العامل المناظرة للعوامل غير المرتبطة حلول مائلة أو غير مباشرة، والمجموعة المناظرة للعوامل المرتبطة تسمى الحل المستقل أو العمودي. ويقدم جدول (13–4) حل غير مباشر لاختبارات وكسلر الفرعية، والعوامل المناظرة لهذه التشبعات لها ارتباط = 0.13 تقريباً.

وفي الحلول غير المباشرة لا تكّون تشبعات العامل ارتباطات بين الاختبارات والعوامل بل تكون مكافئة لأوزان الانحدار. لنرى ماذا يعني هذا، افترض إن اختبار الاستيعاب القرائي له تشبع = 0.81 على العامل الأول وصفر على العامل الثاني.

جدول (13-4) تشبعات العامل لاختبارات وكسلر الفرعية – الحل غير المباشر

المقياس	الاختبار الفرعي	العامل	
		الأول	الثاني
اللفظي	1. المعلومات	0.81	0.00
	2. الفهم	0.81	0.00
	3. الحساب	0.81	0.00
	4. المتشابهات	0.81	0.00
	5. سلاسل الأعداد	0.81	0.00
	6. الألفاظ	0.81	0.00
الأدائي	7. رموز الأعداد	0.00	0.71
	8. إكمال الصورة	0.00	0.71
	9. تصميم المكعبات	0.00	0.71
	10. ترتيب الصور	0.00	0.71
	11. تجميع الأشياء	0.00	0.71

ويشير التشبع صفر إلى انه بعد إن تؤخذ العلاقة الإحصائية بين الاستيعاب والعامل الأول فان الاستيعاب غير مرتبط بالعامل الثاني، وبأسلوب آخر نقول إن الارتباط بين الفهم والعامل الثاني بتثبيت العامل الأول = صفر، ويشير التشبع 0.81 إلى إن الفهم والعامل الأول بينهم علاقة حتى عند تثبيت العامل الثاني ويجب ملاحظة إن تشبع العامل في الدوران غير المباشر لا يكافئ الارتباط الجزئي، بل يكافئ معامل الانحدار المعياري.

ويكون الارتباط بين الاختبارين للعوامل المرتبطة دالة لتشبع العامل والارتباطات بين العوامل ويعبر عن العلاقة لارتباط عاملين على النحو الآتي:

$$\phi\ a_{j1}\ a_{i2} + \phi\ a_{j2}\ a_{i1} + a_{j2}\ a_{i2} + a_{j1}\ a_{i1} = \rho_{ij} \qquad (14\text{-}3)......$$

حيث ترمز ϕ إلى الارتباط بين العاملين ، وصيغة مماثلة لهذه تكون للعوامل الأكثر من اثنين.

عدد العوامل:

يعد تحديد عدد العوامل خطوة مهمة في التحليل العاملي ، وقد يكون عدد العوامل اللازم

لتفسير تجمعات القياس لاي رقم بين 1 و N-1، حيث N هي عدد القياسات . وكمبدأ يمكن تحديد عدد العوامل من علاقة الارتباطات بين أزواج الاختبارات الى تشبعات الاختبارات بهذه العوامل. فعلى سبيل المثال إن كان هنالك عاملين يقاسان بالاختبارات فمن الممكن تحديد تشبعات كل اختبار بالعاملين، وبالتالي فإن المعادلة (13-2) تكون كافية للغرض . ويحقق الارتباط بين الاختبارين i و j وتشبعاتها بهذين العاملين المعادلة:

$$a_{j2} \quad a_{i2} + \; a_{j1} \quad a_{i} \; = \rho_{ij}$$

وان كان هنالك عوامل عددها m فإن الارتباط بين i و j وتشبعات الاختبارين بالعوامل m تحقق المعادلة:

$$a_{jk} \quad a_{ik} \sum_{l=i}^{m} \quad = \rho_{ij}$$

لتقديم مثال محسوس افترض مصفوفة الارتباطات الداخلية المبينة في جدول (13-5) لاختبارات افتراضية أربعة، فما عدد العوامل اللازم لتحقيق المعادلة (13-2) لهذه المصفوفة ؟ الجواب هو 1 وذلك لان $0.8 = a_{11}$ و $0.7 = a_{21}$ و $0.6 = a_{31}$ و $0.5 = a_{41}$ وهي تشبعات العامل، ويمكن التعبير عن أي ارتباط في المصفوفة على النحو الآتي:

$$a_{j1} \quad a_{i1} = a_{ij}$$

كمثال: لـ $0.56 = \rho_{12}$, $0.8 = a_{11}$, $0.7 = a_{21}$

ونحصل على:

$$0.56 = (0.7)\, 0.8 = a_{21} \;\; a_{11} = \rho_{12}$$

جدول (13-5) : مصفوفة الارتباط الافتراضي لاختبارات أربعة

الاختبار	1	2	3	4
1	1.00	-	-	-
2	0.56	1.00	-	-
3	0.48	0.42	1.00	-
4	0.40	0.35	0.30	1.00

لذلك فإن الارتباطات في جدول (13–5) تشير إلى عامل واحد يقاس بالاختبارات الأربعة، حيث إن المعادلة (13–2) يمكن تحقيقها بالتشبعات على عامل واحد. افترض ان ρ_{14}= 0.8 لا 0.4 كما في جدول (13–5)، ففي هذه الحالة نحتاج لأكثر من عامل لتحقيق المعادلة (13–2) بالتشبعات الأربعة المبينة سابقاً وتتحقق المعادلة للارتباطات جميعها ما عدا الارتباط الجديد ρ_{14} ولهذا الارتباط فإن:

$$(0.5)\ 0.8 = a_{41}\ a_{11} \neq 0.80 = \rho_{14}$$

لذلك فلو كانت P_{14}= 0.8 بدلاً من 0.4 فإن عامل واحد لا يفسرالارتباطات عبر المتغيرات الاربعة .

ومن المهم ادراك ان المعادلات المعروضة تربط بين معاملات الارتباط بتشبعات العوامل من خلال بارامترات المجتمع. فالمعادلة (13–2) تتحقق مثلاً عندما يكون عدد العوامل أقل من (1-N) فيما لو عرفنا قيمة الارتباطات في المجتمع. وللارتباطات المحسوبة من العينة تتحقق المعادلة (13–2) بالضبط عندما يكون عدد العوامل = (1-N) فقط . فهل هذا يعني ان ناخذ عوامل عددها يساوي (1–N) في تفسير الارتباطات بين الاختبارات ؟ معظم علماء القياس النفسي لا يأخذون هذه الملاحظة بعين الاعتبار، بل انهم يبرهنون فشل ارتباطات العينة في تحقيق المعادلة (13–2) لعوامل عددها أقل من (1-N) والذي يعكس خطأ المعاينة في ارتباطات العينة . ولبيانات العينة فإنه يمكن التفكير بهذه المشكلة على أنها مشكلة تحديد عدد العوامل اللازم لتحقيق المعادلة (13–2) على وجه التقريب. ودرجة عدم تحقيقها للمعادلة يكون قليلاً بحيث يمكن أن يعزى إلى خطأ المعاينة

نموذج التحليل العاملي:

تعبر المعادلات (13–2) و (13–4) وتعميماتها عن الارتباط بين الاختبارات على أنها دالة لتشبعات العامل، وفي حالة العوامل المرتبطة تعبر عن ارتباطات العوامل. ويمكن اشتقاق هذه المعادلات من نموذج التحليل العاملي:

$$Z_i = \sum_{1=i}^{m} f_k\ a_{ik} + \mu_i \quad \text{.........(5–13)}$$

حيث تمثل Z_i الدرجة الزائية على الاختبار i

و a_{ik} تشبعات الاختبار i على العامل k

و f_k الدرجة على العامل المشترك K

و μ_i الدرجات على العامل الخاص باختبار i

ولغاية الآن لم نعرض التمييز بين العوامل المشتركة والعوامل الخاصة. وقد أشرنا في السابق إلى العوامل المشتركة ببساطة باسم عوامل. والعامل المشترك هو العامل الذي يساهم بالارتباط الملاحظ بين اختبارين أو أكثر من هذه الاختبارات، والعامل الخاص هو العامل الذي يرتبط باختبار واحد فقط. ويضم هذا التعريف إلى الافتراض الذي ينص على إن العامل الخاص غير مرتبط بالعوامل المشتركة جميعا، وأن العوامل الخاصة للاختبارات المختلفة غير مرتبطة ببعضها البعض، فهذا يؤشر إلى إن العوامل الخاصة لا تسهم في الارتباط بين الاختبارات.

الطائفية والتفردية: Communality and uniqueness

سؤال مهم في التحليل العاملي يطرح حول الجزء من تباين الاختبار الذي يرتبط بالتباين على العوامل المشتركة ـ والتي تسمى الطائفية أو التباين الطائفي ويحسب بـ:

$$h^2_i = \sum_{l=k}^{m} a^2_{i}k \qquad (6\text{-}13)$$

للعوامل غير المرتبطة. وللاختبار الفرعي في المعلومات فإن الطائفية تكون:

$$h^2_i = {}^2(0.8) + {}^2(0.1) = 0.65 \qquad (7\text{-}13)$$

والطائفية هي عدد أقل من (1) عندما تستخدم الدرجات الزائية في التعبير عن درجة الاختبار في التحليل العاملي.

كذلك فإن تباين الدرجات الزائية = 1. والطائفية هي الجزء من تباين درجة الاختبار المرتبط بالتباين على العوامل المشتركة. والصيغة الطائفية لعاملين طائفيين مرتبطين تكون:

$$h^2_i = a^2_{i}1 + a^2_{i2} + a_{i1}.a_{i2}\ .\Phi \qquad (8\text{-}13)$$

وصيغة مشابهة تستخدم في حالة أكثر من اختبارين.

وتسمى النسبة من تباين الاختبارات المرتبطة بالتباين على العامل الأول الخصوصية أو التباين الخصوصي، ويحسب بـ:

$$u^2_i = 1 - h^2_i \qquad (9\text{-}13)$$

وللاختبار الفرعي في المعلومات، فإن $u^2_i = 0.35$.ونظرياً يمكن تجزئة التباين الخصوصي إلى عنصرين هما: التباين الخاص و تباين الخطأ. فالتباين الخاص هو الجزء من تباين درجة الاختبار الحقيقية غير المرتبط بتباين الدرجات الحقيقية في أي من الاختبارات الأخرى. دعنا نرمز إلى التباين الخاص بالاختبار في بالرمز S^2_i ولتباين الخطأ بالرمز e^2_i ، فان التباين الكلي يساوي:

$$0.1 \;\; = e^2_i + S^2_i + \; h^2_i$$

وتباين الثبات للاختبار هو $h^2_i + S^2_i$، إذ إن مجموعها يقع بين صفر وواحد ، ويمكن التفكير به على انه نسبة ، لذا فإن $h^2_i + S^2_i$ هي النسبة من التباين الكلي الثابتة أو المكافئة لثبات درجات الاختبار. وكنتيجة يمكن اعتبار h^2_i على أنها الحد الادني لثبات درجات الاختبار i.

مثال بياناته حقيقية

يوضح المثال الآتي تطبيق التحليل العاملي لدراسة استكشافية لصدق البناء لمقاييس فرعية منفصلة لاختبار ذكاء واسع الانتشار. وقد دون كل من جيورتن وبايلي **(Guertin & Bailey 1970)** مصفوفة ارتباطات المقاييس الفرعية الأحد عشر لمقياس وكسلر للبالغين (wais) وتصف البيانات المستخدمة في حساب الارتباطات على أداء (200) فرد من رجال الأمن والاطفاء التي جمعها ما ترازو وزملاءه **(Matarazzo,1964)**، واصغر حجم للعينة في التحليل العاملي وبناءاً على النتائج التجريبية هو التطبيق على (100) مفحوص أو بواقع عشرة أضعاف عدد المتغيرات. وحجم العينة في المثال يتجاوز القيمة المقترحة على أساس التجربة العملية. ويبين جدول (13–6) مصفوفة الارتباطات لهذه.

واحد من القرارات المهمة في أي عملية تحليل عاملي هو عدد العوامل المشار إليها في مصفوفة الارتباطات ويوجد عدد من المحكات لتحديد عدد العوامل. الأول هو عدد قيم التشبعات التي اكبر من1 في مصفوفة الارتباطات، وفي الفقرة التالية نطرح مثالين لتفسير الأساس في اختيار محك قيم التشبعات الأكبر من 1. وقدم موليك **(Mulaik , 1972)** التعريف الرياضي للتشبعات والأساس المنطقي الرياضي لمحك قيمة تشبع اكبر من 1 لعدة العوامل.

ولاختبارات عددها N فإن مصفوفة الارتباطات لها تشبعات عوامل عددها N، ومع أن

طريقة الحصول على قيم التشبعات غير مبينة هنا، فمن المهم ملاحظة أن قيمها النسبية تعتمد على مقدار الارتباطات ونمطها، لذا فمن المنطقي أن قيم التشبعات يشير إلى عدد العوامل، فمثلاً إذا كانت الارتباطات بين أزواج اختبارات وكسلر الفرعية = 0.7، وقيم التشبعات المتبقية = 0.3 ، فمن البديهي القول بان المقاييس الفرعية جميعها تقيس عامل واحد. وإن كان الارتباط بين الاختبارات اللفظية 0.7 وبين المقاييس الأدائية 0.7 والارتباط المتقطع = صفر، فمن البديهي القول إن المقاييس الفرعية تقيس عاملين ولمصفوفة الارتباطات هذه فقد تكون قيم أحد التشبعات 5.2 والأخرى 4.3 والتسعة المتبقية = 0.7 فهذا يشير إلى الحاجة لعاملين لتفسير الارتباطات. ويبين جدول (13–6) قيم التشبعات لارتباطات مقياس وكسلر أن هنالك قيمتين للتشبعات اكبر من 1، مما يؤدي إلى اقتراح وجود عاملين، ومع أن هنالك قيمتين للتشبعات قريبتين من 1، لذا فإنه يوجد أكثر من عاملين، لذا فإن نتائج التحليل العاملي تفترض وجود عاملين أو ثلاثة أو أربعة سيتم تدوينها.

ويبين جدول (13–7) الحل الابتدائي أو غير المدور لعاملين لارتباطات مقياس وكسلر. وتذكر أنه تم حساب تشبعات العامل غير المدور بطريقة مناسبة ثم أجري له تدوير للحصول على تشبعات العامل القابلة للتفسير. وفي الشكل (13–1) تم رسم التشبعات على العامل الابتدائي الأول على المحور الأفقي والتشبعات على العامل الابتدائي الثاني على المحور العمودي. مثال: لتحديد موقع اختبار الفهم الفرعي على الرسم البياني نجد أن احداثياتها تساوي (0.63,-0.34) لاحظ ان الرسم البياني يقترح تجمعين للمتغيرات، يتكون التجمع الأول من المقاييس الفرعية (من 1 إلى 6) والمرسومة في الربع الأسفل إلى اليمين، ويتكون التجمع الثاني من بقية المقاييس من (7-11) وهذه تقع في الربع الأعلى إلى اليمين في الشكل نفسه المقاييس . لاحظ أيضاً ان المحور الأفقي يمر بالمركز مقسماً المقاييس الفرعية الى مجموعتين . وكنتيجة فإن المقاييس الفرعية في كلا المجموعتين لها تشبعات اساسية على العامل الأول ومبينة على المحور الافقي.

جدول (13-6) : جدول الارتباطات عبر اختبارات وكسلر الفرعية :

الاختبار الفرعي	1	2	3	4	5	6	7	8	9	10	11
1. المعلومات	1.00										
2. الفهم	0.37	1.00									
3. الحساب	0.34	0.27	1.00								
4. المتشابهات	0.40	0.25	0.36	1.00							
5. سلاسل الأعداد	0.27	0.38	0.28	0.22	1.00						
6. الألفاظ	0.59	0.46	0.33	0.35	0.29	1.00					
7. الرموز العددية	0.09	0.10	0.18	0.08	0.16	0.08	1.00				
8. إكمال الصور	0.25	0.26	0.32	0.31	0.14	0.27	0.19	1.00			
9. تصميم المكعبات	0.27	0.29	0.38	0.26	0.18	0.24	0.13	0.36	1.00		
10. ترتيب الصور	0.22	0.22	0.29	0.25	0.15	0.28	0.22	0.36	0.30	1.00	
11. تجميع الأشياء	0.26	0.24	0.30	0.20	0.22	0.26	0.17	0.40	0.60	0.25	1.00

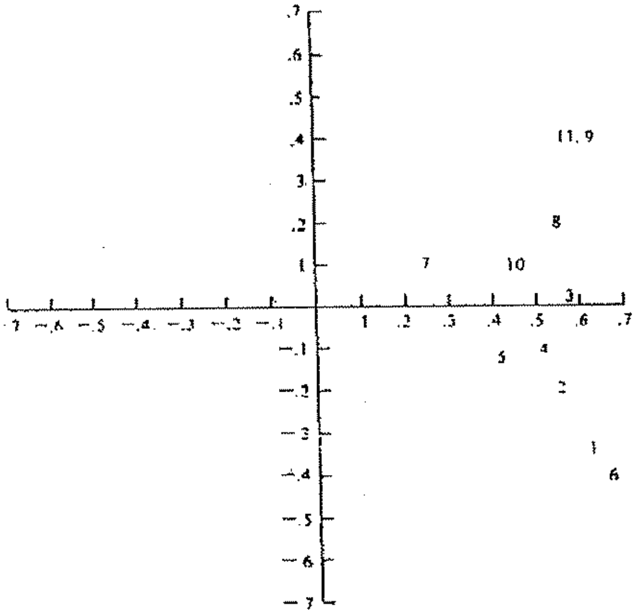
تشبعات مصفوفة الارتباطات هي: 3.79 ، 1.27 ، 0.98, 0.92 ، 0.76 ، 0.71 ، 0.64 ، 0.58 ، 0.55 ، 0.40 0.35 .

W.HG-uer tin, and JD Bailey, بإذن نشر عن:

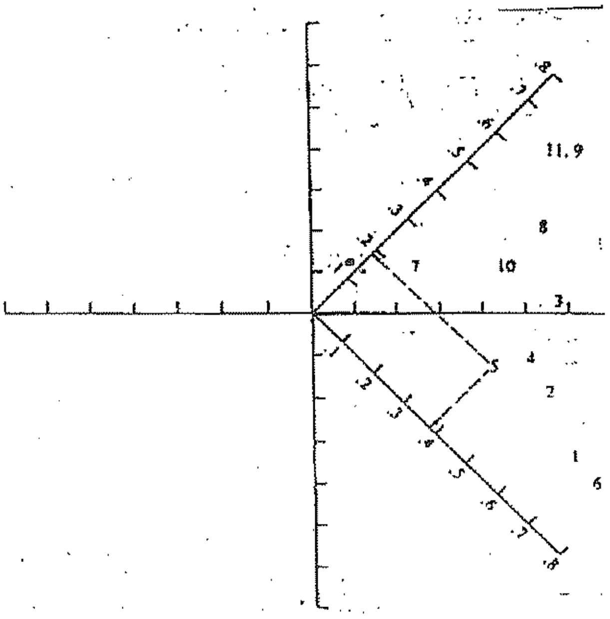
Jr (1970) interoduction to modern factor analysis, Ann arbor, Michi Edward , Brothers.

جدول (13-7) : التحليل العاملي الابتدائي لبيانات اختبار وكسلر

المقياس الفرعي	تشبعات العامل	
	الأول	الثاني
1. المعلومات	0.63	0.34-
2. الاستيعاب	0.55	0.19-
3. الحساب	0.58	0.05
4. المتشابهات	0.51	0.12-
5. سلاسل التعداد	0.42	0.13-
6. الألفاظ	0.66	0.41-
7. رموز الأعداد	0.24	0.12
8. إكمال الصور	0.54	0.20
9. تصميم المكعبات	0.62	0.40
10. ترتيب الصور	0.47	0.11
11. تجميع الأشياء	0.59	0.40



شكل (13-1) : تمثيل بياني لتشبعات العوامل غير المدوّرة لبيانات اختبارات وكسلر



شكل (13-2) : تمثيل بياني للدوران المستقل لبيانات اختبار وكسلر

ويسمح دوران العوامل بايجاد محورين جديدين، احدهما يمر بدرجة اقرب الى المجموعة الاولى والاخر يمر بدرجة اقرب الى المجموعة الثانية، وللحل غيز المباشر فإن المحورين الجديدين يكونان متعامدين . ويبين الشكل (13-2) حل غير مباشر للمحورين الاصليين ، ويوضح الخط المنقط عملية ايجاد تشبعات الاختبارات الفرعية على محوري الدوران وذلك للاختبار الخامس. وكانت التشبعات المدورة على الاختبار الخامس تساوي 0.39 على العامل المدور الأول و 0.19 على العامل المدور الثاني، وبما أن التدوير غير مباشر لذا فان المحاور تبقى متعامدة . وفي مثالنا يمكن ان يعبر كل محور عن مجموعة الاختبارات الفرعية الخاصة به . وبالمقابل فإن تحريك احد المحورين بحيث يصبح اكثر قرباً من مجموعته فإن المحور الاخر سيبتعد عن مجموعته، أما في الدوران المستقل يكون المحورين غير متعامدين لذا فإنه يمكن تعيين المحورين الجديدين بحيث يصبح كل منهما اكثر قرباً من مجموعته.

ويبين الجدول (13-8) تشبعات العوامل للدوران المستقل للعاملين، وللمقارنة فإنه يبين ايضاً التحليل العاملي المدور غير المباشر لكلا العاملين. وقد استخدم الاسلوب الرباعي المباشر للحصول على الدوران المستقل، واستخدم اسلوب تعظيم التباين للحصول على الدوران غير المباشر ، وكان الارتباط بين العاملين للتدوير المستقل (0.55) . وعلى القارئ ان يضع نصب عينيه أن التشبعات الأقل من (0.3) تعد غير مهمة، ويشير الجدول الى ان المتغيرات 1، 2 ، 3 ، 4 ، 5 ، 6 تتشبع بصورة اساسية بالعامل الأول . ويجب تذكر ان استخدام الدوران المستقل يجعل تشبعات العوامل مماثلة لمعاملات الانحدار . لذا فإن التشبعات المنخفضة على العامل الثاني تعني انه عند الأخذ بعين الاعتبار العلاقة بين العامل الاول والمتغيرات (6.5.4.3.2.1) فإنه لا توجد علاقة مع العامل الثاني أو انها تكون ضعيفة . ويشير الجدول (13-8) كذلك الى ان المتغيرات (من 8 إلى11) مشبعة اساساً بالعامل الأول، وتشبع المتغير الثالث متساو لكلا العاملين، وتشبع العامل الرابع مهمل لكلا العاملين .

جدول (13-8) : مجموعتي حل التحليل العاملي الثنائي المستقل وغير المباشر لبيانات اختبار وكسلر

المقياس الفرعي	الطائفية	تشبعات العامل			
		العوامل المستقلة		العوامل غير المباشرة	
		1	2	1	2
1. المعلومات	0.50	0.75	0.06-	0.69	0.18
2. الفهم	0.34	0.54	0.07	0.53	0.23
3. الحساب	0.34	0.29	0.36	0.38	0.43
4. المتشابهات	0.27	0.43	0.14	0.45	0.26
5. سلاسل الاعداد	0.19	0.39	0.07	0.39	0.19
6. الألفاظ	0.60	0.83	0.11-	0.76	0.15
7. رموز الأعداد	0.07	0.02	0.26	0.10	0.25
8. إكمال الصور	0.33	0.10	0.51	0.25	0.52
9. تصميم المكعبات	0.53	0.07-	0.77	0.17	0.71
10. ترتيب الصور	0.22	0.16	0.36	0.27	0.40
11. تجميع الأشياء	0.51	0.08-	0.77	0.16	0.70

وكما هو ملاحظ فإن جدول (13–8) يبين أن تشبعات العامل الناتج عن الدوران غير المباشر باسلوب تعظيم التباين والنمط العام للتشبعات (لكلا التدويرين غير المباشر والمستقل) متشابهة، الا ان تشبعات المتغيرين الثاني والرابع على العامل الثاني كانت اكثر اهمية في التدوير غير المباشر عنه في التدوير المستقل وبصورة مشابهة ، فإن المتغيرين (8 و 10) لها تشبعات اساسية اكثر على العامل الاول في الحل غير المباشر ، لذا فإن الحل المستقل يبسط النمط الذي يعد دليلاً على عدم التحيز في الحل غير المباشر. وهذه مقارنة غير متحيزة بين الحلين المستقل وغير المباشر في التحليل العاملي.

ويتضمن الجدول (13–9) تشبعات العامل التي تم الحصول عليها من عملية التدوير الرباعي المباشر لثلاثة عوامل. العامل الاول يشبه كثيراً العامل الأول في الحل المستقل للعاملين، وتقترح التشبعات على العاملين الاخرين تجزئة العامل الثاني في جدول (13–8) الى عاملين . وتشبع العامل الثاني في جدول (13–9) بالمتغيرين (9 و11) بشكل اساسي، في حين تشبع العامل الثالث بالمتغيرات (7 , 8 , 10) بشكل اساسي . ويقترح تفحص مصفوفة الارتباطات سبباً للتشبع بالعاملين الثاني والثالث، فالارتباط بين المتغيرين (9 و11= 0.60)

وهو اكبر ارتباط في المصفوفة، كذلك فإن الارتباط بين المتغيرين (8 و 10 =0.36) والمتغير 10 لا يوجد له ارتباط اخر عالٍ كهذا، في حين أن المتغير 8 له ارتباطين اخرين عاليين: الاول مع المتغير9 والثاني مع المتغير11. ويبين الجدول (13–10) الحل للعوامل الاربعة الناتجة عن الدوران المستقل، وبصورة اساسية فإن التشبعات على العوامل الثلاثة الاولى مشابهة لما في جدول (13–9). والفرق الرئيسي في تشبعات المتغيرين الخامس والثاني على العامل الأول إذ انخفضت عند الانتقال من الحل بثلاثة عوامل الى اربعة عوامل. وعلى العامل الرابع نرى ان تشبع المتغير الخامس (0.7)، وتشبع المتغير الثاني (0.28)، في حين ان جميع التشبعات الاخرى مهملة. ومن فحص مصفوفة الارتباطات يتبين ان المتغيرين الخامس والثاني ارتباطهما (= 0.38)، وهو الارتباط الاعلى للمتغير الخامس وهذا هو سبب ظهور العامل الرابع.

جدول (13–9) حل العوامل الثلاثة المستقل لبيانات اختباروكسلر

المقياس الفرعي	الطائفية	تشبعات العامل		
		1	2	3
1. المعلومات	0.52	0.75	0.00-	0.05-
2. الفهم	0.34	0.55	0.08	0.01-
3. الحساب	0.34	0.27	0.18	0.28
4. المتشابهات	0.29	0.39	0.04-	0.25
5. سلاسل الاعداد	0.20	0.40	0.06-	0.03
6. الالفاظ	0.60	0.82	0.06-	0.04-
7. رموز الأعداد	0.11	0.4-	0.01	0.36
8. إكمال الصور	0.38	0.06	0.19	0.45
9. تصميم المكعبات	0.57	0.3	0.71	0.05
10. ترتيب الصور	0.35	0.6	0.02-	0.57
11. تجميع الاشياء	0.64	0.00-	0.82	0.03-

وسيركز تفسير التحليل العاملي على الحل باربعة عوامل. العامل الأول مشبع بصورة اساسية بالمعلومات فالمقياس الفرعي (المعلومات) هو اختبارفي المعرفة يحاول تجنب المعرفة الاكاديمية المتخصصة، وصمم اختبار الفهم لقياس الحكم التطبيقي. وعبر الاشياء الاخرى يطلب من المفحوص ان يفسر لماذا تطبيقات معينة. وماذا يمكن عمله ضمن ظروف معينة. واختبار المتشابهات يختبر القدرة على تحديد الطريقة التي يتشابه فيها شيئين اثنين، ويتطلب الاختبار اللفظي (كما هو مبين من اسمه) من المفحوصين بيان معان الكلمات . والى حد معين

فإن فقرات الاختبارات الفرعية الاربعة يمكن الاجابة عليها بناءاً على المعرفة المكتسبة، اي انها لا تتطلب بالضرورة أية قدرة استدلالية. مع أن هذا ليس صحيحاً بعض الشيء لاختباري الفهم والمتشابهات، على سبيل المثال من الممكن في اختبار المتشابهات ان المفحوص تعلم كيف يتشابه الشيئين في الفقرة ، لذا فإنه سيجيب على الفقرة على اساس المعرفة المكتسبة، كذلك فقد لا يكون المفحوص تعلم طبيعة التشابه في فقرة اخرى، ويمكن ان يصل الى الجواب عن طريق الاستدلال اللفظي . لذا يبدو أن العامل الاول يعكس المعرفة اللفظية والاستدلال اللفظي. والعامل الثاني مشبع بصورة اساسية بالاختبارات الفرعية: توضع تصميم المكعبات وتجميع

جدول (13-10) حل العوامل الاربعة المستقل لبيانات مقياس وكسلر

المقياس الفرعي	الطائفية	تشبعات العوامل			
		1	2	1	2
1. المعلومات	0.57	0.76	0.03	0.06-	0.02
2. الاستيعاب	0.36	0.38	0.10	0.01	0.28
3. الحساب	0.34	0.23	0.18	0.26	0.08
4. المتشابهات	0.30	0.42	0.01	0.22	0.02-
5. سلاسل التعداد	0.57	0.09	0.02	0.04	0.70
6. الألفاظ	0.60	0.78	0.02-	0.02-	0.06
7. رموز الأعداد	0.15	0.08-	0.01	0.37	0.11
8. إكمال الصور	0.40	0.13	0.23	0.43	0.11-
9. تصميم المكعبات	0.61	0.03	0.79	0.00-	0.02-
10. ترتيب الصور	0.35	0.14	0.02	0.52	0.07-
11. تجميع الأشياء	0.60	0.04-	0.79	0.02-	0.03

الاشياء، ففي **الاول** يستخدم المفحوص مكعبات ذات جهات لونها احمر واخرى لونها ابيض ليبني اشياء تشبه تصميم شكل مرسوم على بطاقة، وفي الاخر يضع المفحوص القطع الخشبية التي تؤلف شيء ما.

ويبدو ان **العامل الثاني** يعكس القدرة على عمل شيء كلي من اجزائه المؤلفة له. ففي تصميم المكعبات يعرف المفحوص الشكل العام للشيء وعليه ان يجمع الاجزاء لتكوينه. وفي تجميع الاشياء يجمع المفحوص الاشياء التي تؤلف شيء ما كما هو في الطبيعة ومحاولة تخيل الكل من اجزاؤه. واقترحت انستازي (Anastasi,1968) ان هذا العامل يتكون من الجمع بين سرعة الادراك والتخيل المكاني وهذه العوامل تكون منفصلة في دراسات اخرى .

العامل الثالث مشبع بصورة اساسية باكمال الصور وترتيب الصور ولمدى أقل بالرموز العددية، ففي اختبار تكملة الصور يطلب من المفحوص ذكر الاشياء الناقصة في الصورة ، ويتطلب ترتيب الصور من المفحوص وضع سلسلة صور في ترتيب يتحدث عن قصة . وكلا الاختبارين يتطلب علاقة الجزء بالكل ، ففي ترتيب الصور على المفحوص ان يحدد مفاهيمياً ماذا يحدث في كل صورة ثم يميز أو يكتشف العلاقة التي تربط الصور معاً وفي تكملة الصور على المفحوص ان يميز الصورة الكلية ثم يكتشف الجزء المفقود . كذلك تستخدم اختبارات تجميع الاشياء وتصميم المكعبات علاقة الجزء بالكل ايضاً، ومن الممكن ان الاختبارات الفرعية الاربعة تتطلب قدرات متشابهة فاختبارات الاشياء تتطلب قدرات لا تتطلبها اختبارات الصور والعكس صحيح .

العامل الرابع مشبع بصورة اساسية بسلاسل الاعداد التي تتطلب من المفحوص ان يتعلم قائمة اعداد ثم يعيد القائمة باتجاه امامي او عكسي ويبدو أن هذا عامل ذاكرة غير مستخدم باختبار اخر عدا الفهم.

ان اختيار حل العوامل الاربعة يفسر على انه قرار بيداغوجي، وعملياً فإن العديد من مستخدمي التحليل العاملي لا يختارون هذا الحل ، لان العامل الرابع يتحدد بدرجة كبيرة بمتغير واحد. فسلاسل الاعداد والتشبع باختبار الفهم، من الصعب تفسيره، علاوة على ان تفسير العامل الاول للمعلومات والاستدلال اللفظي يجعل من البداهة تشبع اختبار الفهم ظاهر في العامل الأول، إذ ان هذا يظهر في حلول كلا من العاملين والثلاثة. ويجد العديد من جماعة التحليل العاملي ان اي منهما (العاملين أو الثلاثة) مفضلاً على العوامل الأربعة، وفي الاختيار بين العاملين والثلاثة عوامل يبدو أن محك الادراك يبدو مهماً، وتذكر انه بالانتقال من العاملين إلى الثلاثة فان العامل الثاني (أي الادائي) يتجزأ الى عاملين جديدين، فالعامل الثاني الجديد مشبع بصورة اساسية بتصميم المكعبات وتجميع الاشياء، والعامل الثالث مشبع بصورة اساسية بتكملة الصورة وترتيب الصور. ويبدو أن هذا منطقياً إذ يقترح أن اختبارات الاداء التي تستخدم المعالجة تتطلب قدرات مختلفة عن تلك التي تستخدم فحص الصور. بالإضافة الى ان محك الادراك ومحك التركيب البسيط تعد مهمة احياناً في اختيار عدد العوامل التي تفسر، مع ذلك لم يحدث هذا في بيانات بعض مجموعات المثال الحالي، أذ ان محك التركيب البسيط كان تحقيقه واضحاً جداً في عدد من العوامل. ومع عدد أقل من العوامل، فإن العوامل فإن التركيب البسيط لم يتحقق حتى على وجه التقريب ولعدد أكبر من

العوامل لم تكن مدركة أو انها كانت محددة بعدد قليل جداً من المتغيرات. واخيراً فإن المحك القطعي لعدد من العوامل المفسرة هو تكرارها، فعندما تفحص المتغيرات نفسها في دراسات عديدة، فإن العوامل التي تتكرر في الدراسات هي التي يجب أن تفسر.

بالإضافة الى التحقق من تشبعات العامل فإنه يحب التحقق من الطائفية للاختبار ومقارنتها بثباته. وهدف هذا التحقق والمقارنة هو تحديد:

1/ نسبة تباين الاختبار المرتبط بتباين العامل المشترك كما يؤشر اليه من خلال الطائفية للاختبار.

2/ نسبة تباين الاختبار الثابت ولكنه مرتبط بالعامل الخاص بالاختبار، وتحسب هذه النسبة من الفرق بين ثبات الاختبار وطائفية الاختبار.

وللتوضيح افترض أن الاختبار الفرعي (الاستيعاب) في حالة العوامل الثلاثة على انها الحل كان له تشبع 0.55 على العامل الأول، والذي يقترح انه يقيس المعرفة اللفظية والاستدلال. من جهة اخرى الطائفية له تساوي (0.34)، وهذه تشير الى ان الاستيعاب يشارك بـ34% من التبايـن مع العوامل الطائفية الثلاثة. وقد دون وكسلر (Wechsler, 1558) معامل ارتباط تجزئة سبيرمان - براون المصحح لاختبار الفهم القيمة 0.77 وهذه القيمة محسوبة لعينة مؤلفة من 300 مفحوص في الفئة العمرية من 25 - 34 سنة (وذكرت هنا لاهداف توضيحية فقط). واستخدام شكل الثبات هذا فإن النسبة 33 (= 100[0.77 - 0.34]) من التباين في درجات الاستيعاب ثابت ومرتبط بالعامل الخاص لاختبار الفهم. ومن الواضح بعد ذلك ان تشبع 0.55 على العامل الأول يعد من الخطأ اعتبار اختبار الفهم قياس لعامل المعرفة اللفظية والاستدلال بصورة اساسية من جهة أخرى، فإن تشبع تجميع الاشياء على العامل الثاني=0.82 ، وطائفية = 0.65 ، لذا فإنه يشارك ب 65% من تباينه مع العوامل الطائفية الثلاثة، وب 1% من تباينه ثابت ومرتبط بالعامل الخاص بتجميع الاشياء. وكنتيجة فإن تجميع الاشياء بصورة اساسية يعد مقياس للعامل الثاني. ومثال ثالث، تشبع الاختبار اللفظي على العامل الاول 0.82 وطائفية =0.60، لذا فانه يشارك بـ 60% من تباينه مع العوامل الثلاثة. وله ثبات = 0.95 ، لذا فإن 35% من تباينه ثابت ويعزى الى العامل الخاص بالالفاظ ، لذلك فإنه يقيس وبصورة سائدة العامل الأول ، ويتأثر ايضاً بالعامل غير المرتبط بالمتغيرات الاخرى التي تقيس المعرفة والاستدلال اللفظيين .

ونقطة مهمة متعلقة بالتحليل العاملي هو انه ببساطة يحدد تجمعات المتغيرات، ومحاولات تحديد لماذا تتواجد هذه التجمعات التخمينية، ويجب تصورها على شكل فرضية تحتاج الى

المزيد من الاثبات. ايضاً وبالاضافة الى ما كتب عن التحليل العاملي هنا، على القارئ ادراك ان المفحوصين في مثالنا هم رجال الامن والاطفاء، وهؤلاء لا يمثلون فئة البالغين بشكل عام . ولهذا السبب فان نتائج التحليل العاملي هنا قد تختلف بصورة جوهرية عن نتائج انماط اخرى من المفحوصين.

التحليل العاملي الاستكشافي والتوكيدي:

يمكن وصف الباحثين المستخدمين للتحليل العاملي على متصل استكشافي – توكيدي، فالتحليل العاملي يكون استكشافياً محضاً فيما لو لم يكن لدى الفاحص فرضية ويكتنفه الغموض حول عدد او طبيعة العوامل المقاسة بالاختبار.ويكون التحليل العاملي توكيدي اذا كان لدى الفاحص فرضية ويجرى اختبارات احصائية لها وسيتم تقديم ملخص لطبيعة هذه الاختبارات لاحقا في هذا الجزء.

والقليل من الباحثين استكشافيين بصورة محضة، فمن المألوف ان يكون لدى الباحث نظرية عن عدد العوامل المقاسة بالاختبارات وطبيعتها. فمثلا في مثالنا لمقياس وكسلر للبالغين فقد يفترض الباحث وجود عامل لفظي واخر ادائي، ومن الصحيح ايضا ان هذه الفرضية لم تخضع لاختبارات احصائية في التحليل العاملي النموذجي. فضلا عن ان هذه العوامل تم استخلاصها وتدويرها، وخضعت لبعض الاحكام الشخصية المتعلقة بمدى مطابقة النتائج لفرضية الباحث. وهذا الفشل في اختبار الفرضية يجعل الدراسة استكشافية اكثر مما هي اثباتية توكيدية. وسيتم توضيح مشكلة الاسلوب الاستكشافي في هذا من خلال التحليل العاملي لبيانات وكسلر .فقد يفسر بعض الباحثين الحل من خلال العاملين الاثنين والذي يوفر قياس يدعم فرضية العاملين اللفظي والادائي. وقد يفسر اخرون بالاعتماد على الحل من خلال العوامل الثلاثة والذي يوفر دعم اقل لفرضية العاملين اللفظي والادائي. وفي التحليل العاملي التوكيدي فان الباحث يضع فرضية حول القيم العددية لبعض المعالم (تشبعات العامل، والارتباطات بين العوامل، والخصوصية). والاكثر شيوعاً هو ان يضع الباحث فرضية ان تشبعات عامل معين =صفر، على سبيل المثال فللباحث الذي يتوقع عامل لفظي واخر ادائي لبيانات وكسلر عليه ان يضع فرضية ان الاختبارات الادائية جميعها لها تشبعات = صفر على العامل الاول، وان الاختبارات اللفظية جميعها لهاتشبعات تساوي صفر على العامل الثاني .

ويبين الجدول (13–11) هذه الفرضية تشبعات الاختبارت الادائية الفرعية جميعها على العامل الاول تساوي صفر، في حين ان تشبعات الاختبارات اللفظية الفرعية جميعها على العامل الثاني تساوي صفر، ولم تبين الفرضية القيم العددية لكل تشبع ومرمز لها بالرمز X.

وتحسب هذه التشبعات من البيانات كما هو الحال في حساب المعالم الاخرى (الارتباط بين العوامل والخصوصية) والتي لم تفترض لها قيم معينة والتشبعات الصفرية في جدول (13-11) لم يتم حسابها. فقد وضعت لها القيمة صفر في العملية الحسابية.

جدول (13-11) تشبعات العوامل ضمن شروط فرضية العاملين اللفظي - الادائي:

المقياس	الاختبار الفرعي	العوامل	
		1	2
اللفظي	1. المعلومات	x	صفر
	2. الفهم	x	صفر
	3. الحساب	x	صفر
	4. المتشابهات	x	صفر
	5. سلاسل الاعداد	x	صفر
	6. الالفاظ	x	صفر
الأدائي	7. رموز الأعداد	صفر	x
	8. إكمال الصور	صفر	x
	9. تصميم المكعبات	صفر	x
	10. ترتيب الصور	صفر	x
	11. تجميع الأشياء	صفر	x

وبعد حساب القيم العددية لكل من تشبعات العامل والارتباط بين العوامل ، والخصوصية بتطبيق المعادلات (13-2) و(13-3) او تعميماتها ،فانه يمكن استخدام اي معادلة مناسبة لحساب الارتباط بين الاختبارين، ويعد اختبار القيم الافتراضية مقارنة اساسية ما بين هذه الارتباطات والارتباطات المحسوبة من درجات الاختبار.فالارتباطات الاولى تعكس القيم الافتراضية في حين ان الاخرى لا تعكسها . في حالة كون القيم الافتراضية خاطئة فان مجموعتي الارتباطات تكون مختلفة تماما .ويؤدي هذا الفرق الى رفض القيم الافتراضية. وقد قدم جورزكوج (joreskog,1968) عرضاً عاماً للاسلوب التوكيدي في التحليل العاملي مع امثلة عديدة ومثل هذه التحليلات يمكن اجراؤها باستخدام احد البرامج المتخصصة في التحليل العاملي والبرامج الاكثر شيوعاًلهذا التحليل هو (LISREL V1) Joreskog ,Sorbom

1984. ويمكن الحصول على هذا البرنامج من قبل " scientifis software ,Inc" كما انه مضافاً الى البرنامج الاحصائي SPSS.

الخلاصة:

احد اهداف التحليل العاملي هو تحديد عدد العوامل المشتركة التي تسهم في تفسير نمط الارتباطات بين ازواج الاختبارات المختلفة في مجموعة اختبارات. والعامل المشترك هو متغير كامن او غير ملاحظ ويربط بين درجات اختبارين او اكثر. ويشار الى العلاقة بين درجات الاختبار والدرجات على العامل المشترك بتشبع الاختبار بالعامل المشترك المعين، وعندما لا يوجد ارتباط بين العوامل المشتركة فإن تشبع العامل هو الارتباط بين درجات الاختبار ودرجات العامل المشترك . وعندما يوجد ارتباط بين العوامل المشتركة فإن تشبع العامل يكافئ معامل الانحدار المعياري لانحدار الاختبار على مجموعة العوامل المشتركة. وهدف ثاني للتحليل العاملي هو تحديد طبيعة العوامل المشتركة التي تعزى لها الارتباطات الداخلية للاختبار، وهذا التحديد معقد نوعاً ما ،وذلك لانه يوجد في الحقيقة اكثر من مجموعة واحدة من العوامل المشتركة التي يمكن ان تعزى اليها نمط الارتباطات، وهذه تعد مشكلة عدم محدودية الدوران ،فمشكلة اختيار مجموعة عوامل مشتركة مفسرة تم حلها على اساس استخدام تشبعات العوامل التي تحقق محك تركيب بسيط .وهو ان العوامل المشتركة المفسرة هي عوامل تتصف بتشبعات عامل يحقق محك التركيب البسيط .وعملياً تم حل المشكلة باستخدام احد اساليب الدوران المناسبة . وقد صممت هذه الأساليب لإنتاج تشبعات عامل هي تقريب للتركيب البسيط.

وهدف ثالث للتحليل العاملي هو تحديد نسبة التباين لمتغير ملاحظ ومرتبط بتباين عوامل مشتركة وتسمى هذه النسبة من التباين بالطائفية. وتكملها التفردية وهي نسبة التباين المتغير الملاحظ غير المرتبط بتباين العامل المشترك. ويمكن تبيين طائفية المتغير على انها الحد الادنى لثبات المتغير.

التمارين:

1/ يبن الجدول ادناه معاملات الارتباط عبر ستة موضوعات تدرس في المدارس الاسكتلندية

	الموضوعات المدرسية					
	1	2	3	4	5	6
1	0.100	0.439	0.410	0.288	0.329	0.248
2	0.439	0.1000	0.351	0.354	0.320	0.329
3	0.410	0.351	1.000	0.164	0.190	0.181
4	0.228	0.354	0.164	1.000	0.595	0.470
5	0.329	0.320	0.190	0.595	1.000	0.464
6	0.348	0.329	0.181	0.475	0.464	1.000

أ- بالاعتماد على هذه الارتباطات، ما عدد العوامل التي يبدو انها تساهم في الارتباطات الداخلية؟ وتشبعات العامل المدور بتعظيم التباين للحل ثنائي العامل هو:

	تشبعات العامل	
الموضوع	1	2
اللغة الاسكتلندية	0.229	0.659
الانجليزي	0.323	0.551
التاريخ	0.086	0.591
الحساب	0.771	0.173
الجبر	0.720	0.215
الهندسة	0.577	0.213

ب- بالاعتماد على تشبعات العوامل، ما معامل الارتباط المحسوب بين درجات اللغة الاسكتلندية ودرجات الحساب.

ج- على افتراض ان الاختبارات تقيس عاملين ،فلماذا تعود الفجوة بين الارتباطات المحسوبة والمدونة في مصفوفة الارتباط.

د- ما الطائفية لموضوع الهندسة .

هـ- ما التخاصية (او التفرد) لموضوع الجبر .

و- ما قيمة الحد الادنى المحسوب لدرجات التاريخ؟

ز- على افتراض ان معامل ثبات درجات اللغة الاسكتلندية، الانجليزي، التاريخ 0.75 تقريباً . فهل يمكن اعتبار أي من هذه المتغيرات مقياس أولي (أساسي) للعامل الثاني.

ح- بإعطاءالنتائج المناسبة، كيف يمكنك التحقق مما إذا كانت هنالك عوامل اضافية تلزم لتساهم في معاملات الارتباط.

2/ استخدمت قائمةصفات الكآبة في قياس حالات مختصرة لانماط الاحباط وتتضمن هذه القائمة قوائم صفات طلب من المستجيبين ان يختاروا الصفة التي تصف الطريقة التي يشعرون بها . أجرى روث ولوبين (**Ruth Lubin 1981**). تحليلا عاملياً لسبعة عشر صفة وصححت الاستجابات بالتدريج الثنائي. وان كان مؤشر نمط الكآبة "عدم السعادة على سبيل المثال "واختاره المستجيب فانه يعطى الدرجة (1) وان كان مؤشر النشاط "على سبيل المثال واختاره المستجيب فانه يعطى الدرجة (صفر)، وكذلك استجابات اخرى اعطيت الدرجة (صفر). وكانت تشبعات العوامل لتحليل العامل المدور افقياً على النحو الآتي، مع العلم بان معامل الارتباط المحسوب بين العاملين قيمته (0.21)

قوائم المصنفات	التشبعات	
	عامل 1	عامل 2
عدم السعادة	0.63	0.03
الخمول	0.58	0.07
اليأس	0.56	0.01
مسدل	0.56	0.12
الاحباط	0.51	0.12
الخسارة	0.47	0.04-
اليأس	0.45	0.02
الوحدة	0.42	0.04-
الوهن	0.39	0.02-
الارهاق	0.39	0.48
حزين	0.34	0.10
الحرية	0.03-	0.53
الجودة	0.24	0.48
الامن	0.13	0.49
النشاط	0.08	0.47
رباطة الجأش	0.01-	0.48
الغضب	0.00	0.47

أ- اعط تفسير لتشبعات العوامل.

ب-كيف تتوقع لفرد نمطة مكتئب ان تكون درجته على العامل الاول عالية ام منخفضة نسبياً؟ لماذا؟

ج- فسر معامل الارتباط بين العاملين الذين حددتهما في أ.

الوحدة الرابعة

تحليل الفقرات في تطوير الاختبار

الفصل الرابع عشر

14

تحليل الفقرات

الفصل الرابع عشر

تحليل الفقرات

هدف عام لبناء الإختبارات هو الوصول إلى اختبار بأقل عدد من الفقرات ويؤدي إلى درجات لها درجة مناسبة من الثبات والصدق للاستخدامات الخاصة بالاختبار. ويرافق هذا تجريب ميداني لعدد كبير من الفقرات، واختبار مجموعة فرعية منها تساهم بأكبر قدر من ثبات الاختبار أو صدقه. وفي عملية بناء اختبار جديد (أو اختصار عدد فقرات قيد الاستعمال) فإن المجموعة الأخيرة من الفقرات تحدد عادة خلال عملية تسمى تحليل الفقرات. وتحليل الفقرات مصطلح واسع الاستخدام في تحديد الحسابات واختبار أية خصائص احصائية لاستجابات المفحوصين على فقرة اختبارية. ويتم التحقق من معالم الفقرات والتي تقع ضمن ثلاث فئات عامة هي:

1. معاملات تصف توزيع الاستجابات لفقرة مفردة (مثل المتوسط والتباين للاستجابات على الفقرة).
2. معاملات تصف درجة العلاقة بين الاستجابات على الفقرة وأي محك مناسب.
3. معاملات عبارة عن دوال لكل من تباين الفقرة وعلاقته بالمحك.

وسنقدم في الأجزاء التالية صيغاً تعريفية لبعض المعاملات الشائعة في كل فئة، وبعدها سنعرض بيانات مثال توضح كيف أن المعلومات عن معالم الفقرات تؤخذ بعين الاعتبار في اختيار الفقرات أو مراجعة القرارات في الاختبارات معيارية المرجع. وسيتم تقديم طرق تحليل الاختبارات محكية المرجع في الأجزاء الأخيرة من الفصل.

صعوبة الفقرة، والمتوسط، والتباين:

عندما تكون الفقرة ثنائية التصحيح فإن متوسط الفقرة يتطابق مع نسبة المفحوصين الذين أجابوا على الفقرة إجابة صحيحة، ويرمز لهذه النسبة بالرمز P_i، ويطلق عليه اسم صعوبة الفقرة. ونوقش هذا المعلم ببعض التفصيل في الفصل الخامس. تذكر أن قيمة P_i تتراوح بين صفر إلى 1.00، وكما سترى فيما بعد فإن معظم معالم الفقرة تتأثر بصعوبتها. ويرتبط متوسط الدرجة الكلي مباشرة بصعوبة الفقرة لأن :

$$\mu_x = \sum_i P_i \quad \text{.............(1-14)}$$

علاوة على أنه عند اهتمامنا بمستوى متوسط الفقرة في الاختبار، فإنه يمكن الحصول عليها من :

$$\mu_p = (\mu_x) / K \quad \text{.............(2-14)}$$

حيث ترمز K إلى عدد فقرات الاختبار. وعند وصف مجموعات المفحوصين على اختبارات عدة أطوالها مختلفة، فإن متوسط صعوبة الفقرة قد يفضل على متوسطات الدرجة الخام (والذي يتغير كدالة لطول الاختبار). وكما بين الفصل الخامس فإن مستوى صعوبة الفقرة يضبط بتباين الفقرة، لأن:

$$\sigma_i^2 = qi\ Pi$$

كذلك تم تأسيس – وعلى افتراض درجة ثابتة من الارتباط عبر الفقرات – أن تباين الدرجة الكلي يصل إلى أعلى قيمة عندما تكون $P_i = 0.5$. لذا فإنه من المدهش أن نتعلم أنه ولمعظم اختبارات التحصيل والاستعداد المنشورة المصممة لتفسير درجاتها على أنها معيارية المرجع، أن صعوبة فقراتها تقع في الفترة 0.60–0.80 ويعود سبب هذا إلى صياغة الفقرات المستخدمة في مثل هذه الاختبارات.

ولفهم كيفية تأثير صياغة الفقرة على قيمة P مثل :

$28 \times 7 =$ ....

من الواضح أن فرصة التخمين للمفحوص الذي لا يعرف الإجابة الصحيحة قليلة، لذا فإن قيمة Pالملاحظة وبالتالي تباينها تعد دلالة لمعرفة المفحوصين (أي الدرجات الحقيقية على الفقرة) افترض الآن أن الفقرة نفسها كانت صياغتها على النحو الآتي:

$28 \times 7 =$ ......

أ – 186

ب– 196

جـ– 287

د – 554

لأن هذه الصياغة تسمح للمفحوصين بوضع إشارة على الإجابة الصحيحة بالتخمين فإن قيمة أ الملاحظة تتأثر بكل من درجات المفحوصين الحقيقية على الفقرة والتخمين. وضمن

افتراض عشوائية التخمين فإن قيمة أ الملاحظة يتوقع أنها مجموع نسبة المفحوصين الذين يعرفون الإجابة الصحيحة (قيمة أ الحقيقية) و $\frac{1}{m}$ من نسبة الذين لا يعرفون الإجابة الصحيحة، حيث ترمز m إلى عدد الإختيارات أو بدائل الإجابة. ولتعظيم تباين الدرجة الحقيقية الكلي للاختبار فمن الضروري تعظيم تباينات الدرجة الحقيقية للفقرة. ويمكن تعظيم التباين عندما يجيب نصف المفحوصين على الفقرة اعتماداً على معرفتهم بينما النصف الآخر لا يستطيع ذلك، وعبر هذا النصف الذي لا يعرف الإجابة الصحيحة فإن لكل مفحوص احتمالية = $\frac{1}{m}$ ليجيب على الفقرة إجابة صحيحة بالتخمين العشوائي من خلال البدائل التي عددها m. لذا فان نسبة المفحوصين المتوقع ان يجيبو اجابة صحيحة بالتخمين = $\frac{0.5}{m}$ بالتالي فإن الصعوبة الملاحظة للفقرة بأقصى تباين درجة حقيقية يتوقع أن يساوي:

$$\frac{0.50}{m} + \quad 0.50 \ = \ P_o$$

جدول (14-1) توزيع الاستجابات وقيمP على فقرات موضوعية لها احتمالات استجابة مختلفة

po للورد	po	نسبة الذين ضمنوا الاجابة	نسبة الذين يعرفون الإجابة	عدد بدائل الإجابة
0.74	$0.62 = \frac{0.50}{4} + 0.50$	$\frac{0.50}{4}$	0.50	4
0.77	$0.67 = \frac{0.50}{3} + 0.50$	$\frac{0.50}{3}$	0.50	3
0.85	$0.85 = \frac{0.50}{2} + 0.50$	$\frac{0.50}{2}$	0.50	2

ويؤسس جدول (14-1) كيفية حساب قيم P_o لفقرات ذات عدد مختلف من البدائل. والقيم المحسوبة في هذا الجدول أقل نوعاً من قيمة P المثالية للاختبارات معيارية المرجع. وأثبت لورد (Lord, 1952) في دراسة محاكاة ان الثبات يتحسن عند اختيار فقرات لها قيم P أكبر من تلك المحسوبة بالتعديل الناجم عن التخمين وقيم P التي أوصى بها لورد مبينة في جدول (14-1). ومن جهة التطبيق العملي فإن قيم لورد تعتمد على بيانات استجابة محاكاة (افتراضية وليست واقعية)، مع ذلك فإنها تبقى أكثر معقولية من تلك المحسوبة بنموذج التخمين العشوائي، لأن العديد من المفحوصين لديهم معرفة جزئية تمكنهم من حذف واحد أو أكثر من البدائل الخاطئة قبل إجراء عملية التخمين، مما يجعل احتمالية الإجابة الصحيحة أكبر من النسبة $\frac{1}{m}$.

تمييز الفقرات:

تهدف العديد من الاختبارات تقديم معلومات عن الفروق الفردية سواء على المقياس (الاختبار) أو بناء على محك خارجي يفترض أن تتنبأ به درجات الاختبار. وفي كلا الحالتين فإن المعلم الذي نهتم به في اختبار الفقرات يجب أن يكون مؤشراً لفعالية تمييز الفقرة بين المفحوصين ذوي الأداء العالي وذوي الأداء المنخفض على المحك نفسه. ولا يوجد في بعض الحالات مقياس أنسب للبناء من الدرجة الكلية على الاختبار نفسه (الاختبار التحصيلي الصفي هو مثال أساسي لهذا الموقف). وفي هذه الحالة تستخدم الدرجة الكلية كتعريف اجرائي للموقف النسبي للمفحوص على البناء قيد الاهتمام. وفي هذا المحك الداخلي يكون الهدف تحديد الفقرات التي تكون احتمالية الإجابة الصحيحة عليها عالية من قبل فئة التحصيل العالي. ومنخفضة في الوقت نفسه من قبل فئة التحصيل المتدني وفي الاختبارات التحصيلية على سبيل المثال نقول أن الفقرة تميز أو تفاضل بين المفحوصين الذين يعرفون المادة والذين لا يعرفونها من جهة أخرى فإننا نشك في الفقرة التي يتساوى فيها نجاح الفئتين. فمثل هذه الفقرات يبدو أنها لا تقيس البناء نفسه الذي تقيسه بقية الفقرات. ومن غير المرغوب به أيضاً أن يكون هنالك فقرات تفشل الفئة العليا في الإجابة عليها إجابة صحيحة وتنجح الفئة الدنيا في الإجابة عليها بصورة صحيحة، فمثل هذه الفقرات يكون تمييزها سالب.

وتستخدم خمسة معالم على أنها مؤشرات لفعالية تمييز الفقرة، وسيتم وصفها في الجزء التالي. يعتمد أحدها على التفاضل بين مجموعات من المفحوصين تم تحديدهم من خلال درجات قطع عند نقاط محددة على توزيع درجات المحك، والأربعة المتبقية هي أنواع مختلفة من معاملات الارتباط. ومن المهم تمييز أن هذه المعالم جميعها يمكن تطبيقها بالتساوي على المواقف الذي ترتبط درجة الفقرة فيه بدرجة الأداء الكلي أو في الموقف الذي ترتبط فيه درجة الفقرة بالأداء على محك خارجي. ومع أن التطبيق الأخير مثيراً للجدل والخلاف نوعاً ما لأسباب سيتم بحثها فيما بعد، إلا أن اختيار محك داخلي (الدرجة الكلية) لا يغير من الصيغ المحددة أو الطرائق الحسابية.

دليل التمييز : Index of Discrimination

هو معلم تمييز بسيط يطلق عليه اسم دليل التمييز يمكن تطبيقه فقط على الفقرات ثنائية التصحيح ويتطلب حسابة تحديد نقطة أو اثنتان على توزيع درجات المحك على أنها درجات القطع، وتفصل المفحوصين إلى مجموعات درجاتهم أعلى أو أدنى من درجات القطع هذه. على

سبيل المثال، إن كان اهتمام مطور الاختبار في اختيار الفقرات التي تميز بين المفحوصين بناءاً على محك داخلي (درجة الاختبار الكلية)، والمجموعات تتألف من أعلى 50% وأدنى50% من المفحوصين اعتماداً على درجاتهم الكلية، وكانت مجموعة المفحوصين كبيرة فليس من الضروري استخدام الفئة العليا أو الدنيا جميعها. وأجرى كيلي (**Kelley , 1939**) دراسة تقليدية تحت شروط معينة، ووجد أنه يمكن الحصول على معامل تمييز فقرة أكثر حساسية واستقراراً عند استخدام أعلى 27 % وأدنى 27 % من المفحوصين وذلك عندما يكون حجم العينة كبيراً. وفي الحقيقة يمكن الحصول على النتائج نفسها من خلال أعلى وأدنى 30 % أو 50% من المفحوصين (Benchert & Mendoza , 1979 ; Englehart , 1965) وحالما يتم تحديد الفئتين العليا والدنيا، فإن دليل التمييز يحسب بـ :

$$P_L - P_a = D \quad .... (4\text{–}14)$$

حيث ترمز Pa إلى نسبة الذين إجابو على الفقرة إجابة صحيحة من الفئة العليا، و P_L هي النسبة نفسها للفئة الدنيا. وتتراوح قيم D بين – 1.00 إلى 1.00). وتشير القيم الموجبة إلى تمييز للفقرة يحابي الفقرة العليا من المفحوصين، وتشير القيم السالبة إلى تمييز عكسي للفقرة يحابي الفئة الدنيا.

جدول (14–2): انماط استجابة توضيحية لعشرة مفحوصين على فقرات ثلاثة مختارة .

	المفحوصين									
الفقرة	1+	2+	3-	4+	5	6	7	8-	9	10-
1	0	1	1	1	0	1	1	0	1	0
2	1	1	1	1	0	0	0	1	0	1
3	1	1	1	1	0	1	1	1	1	1
.										
الدرجة الكلية على 30 فقرة	25	27	15	24	20	18	16	14	22	10

ويبين جدول (14–2) استجابات 10 مفحوصين على 3 فقرات مختارة من اختبار مؤلف من **30** فقرة ، ومن الطبيعي أنه لتحليل الفقرات يلزمنا عينة أكبر، واستخدم هذا المثال فقط من باب التوضيح. وقد وضعت إشارة + لتمييز فئة المفحوصين العليا (أعلى 30 %) وإشارة

- لتمييز الفئة الدنيا (أدنى 30%) واستخدمت استجابات أعلى ثلاثة مفحوصين وأدنى ثلاثة منهم لحساب D للفقرات (1، 2 ، 3) في الجدول لاحظ أن الفقرات 1 ، 2 لها قيمة P نفسها (0.60)، إلا أن قيم D لها تختلف بدلالة. وقد قدم ابيل (Ebel, 1965) ارشادات لتفسير قيم D اعتماداً على خبرته العملية، وذلك عندما يتم اختيار المجموعات بناءاً على محك الدرجة الكلية، وعلى النحو الآتي :

1. إذا كانت $D \leq 0.40$ ، فإن الفقرة تلبي الغرض أو الهدف.
2. إذا كانت $30 \leq D \geq 0.39$ ، فإن الفقرة تتطلب مراجعة قليلة.
3. إذا كانت $0.20 < D \geq 0.29$ ، فإن الفقرة تقع على الحد الفاصل وتحتاج إلى مراجعة.
4. إذا كانت $D \geq 19$، فإنه يجب حذف الفقرة أو إجراء مراجعة تامة لها.

وأحد محاذير استخدام D أن ليس لها توزيع معاينة معروف أو متفق عليه ،فمن غير الممكن الإجابة عن أسئلة مثل ما هي قيمة D وبدلالة عن الصفر؟ أو ما هو مقدار الفرق بين قيم D الدال احصائيا؟ وعلى الرغم من ذلك ولسهولة حساب D وتفسيرها فإن دليل التمييز يبقى واحداً من أكثر الطرق شيوعاً في تدوين فاعلية تمييز الفقرة، وهو مناسب وعلى وجه الخصوص في تحليل فقرات الاختبارات الصفية.

المعاملات الارتباطية للتمييز:

ناقشنا في الفصل الثاني معامل ارتباط بيرسون على أنه مقياس لدرجة العلاقة الخطية بين متغيرين، فإن كان مدى درجات فقرات الاختبار من (1 إلى 4) أو (1 إلى 5) أو أكثر من ذلك (مثل فقرات استبانة اتجاهات ليكرت، فإنه يمكن استخدام هذه الصيغة لتقدير درجة العلاقة بين درجات الفقرة ودرجات المحك. وعلى الرغم من امكانية استخدام صيغة الارتباط نفسها للفقرات ثنائية التصحيح، فقد تم تطوير صيغ خاصة أسهل استخداماً في الحسابات اليدوية عندما تكون واحدة أو كلا المتغيرين ثنائية الدرجة. وسيتم وصف معاملات ارتباط أربعة في الجزء التالي:

جدول (14-3): حسابات توضيحية لمعاملات التمييز لفقرات جدول (14-2) ، حيث $\mu = 19.10$ ، و $\sigma_x = 5.17$

الفقرة	P	$\mu+$	دليل التمييز $Pl - Pu = D$	معامل ارتباط بونيت بايسيريال $q/p \frac{\mu-\mu+}{\sigma} = rpbis$	معامل ارتباط بايسيريال $(y/p) \frac{\mu-\mu+}{\sigma} = rbis$
1	0.60	20.33	0.34 =0.33 -0.67	$0.29 = (1.22) \frac{19.10-20.33}{5.17}$	$(.3867/ .60) \frac{19.10-20.33}{5.17}$ 0.37=
2	0.60	19.17	0 =1.00 -1.00	$.016 = (1.22) \frac{19.10-19.17}{5.17}$	$(.3867/ .60) \frac{19.10-19.17}{5.17}$ 0.021=
3	0.60	19.00	0 =1.00 -1.00	$0.06 = (1.22) \frac{19.10-19.00}{5.17}$	$(.3867/ .60) \frac{19.10-19.00}{5.17}$ 0.10=

معامل ارتباط بونيت بايسيريال:

عندما يكون مطور الاختبار مهتماً بمعرفة مدى الارتباط بين الأداء على فقرة ثنائية التصحيح (درجاتها صفر، 1) ودرجة الاختبار الكلية (أو أي محك آخر تدريجه متصل)، نستخدم هنا صيغة حسابية بسيطة لمعامل ارتباط بيرسون يسمى معامل ارتباط بوينت بايسيريال، ويعرف على النحو التالي:

$$P_{pbis} = \frac{\mu_+ - \mu_x}{\sigma_x} \sqrt{q / P} \qquad (5\text{-}14)............$$

حيث يرمز μ_+ إلى متوسط درجة المحك للذين أجابوا على الفقرة إجابة صحيحة، وμ_x إلى متوسط درجة المحك للمجموعة الكلية، وσx إلى الإنحراف المعياري، و P إلى صعوبة الفقرة، و $q = 1 - p$. ويبين جدول (14–3) حساب معاملات ارتباط بونيت بايسيريال بين الدرجة الكلية ودرجات الفقرة وذلك للفقرات المدونة في جدول (14–2).

وتكون قيمة معامل ارتباط بونيت بايسيريال نوعاً ما أكبر من القيمة الحقيقية، وذلك لأن درجة الفقرة لها مساهمة في الدرجة الكلية لكل مفحوص. فإن كان عدد الفقرات كبير نسبياً

(25 أو أكثر) فإن هذه الحقيقة نادراً ما تؤدي إلى مشكلة، وعليك تصحيح الصيغة بحل المشكلة للعدد القليل من الفقرات باستخدام الصيغة:

$$\frac{\sigma_i - \sigma_x \quad \sigma_{xi}}{\sigma_i\, \sigma_x\, P_{xi}\, 2 - \ \sigma_x^{2+}\sigma_\iota^{2}} = \rho 1(x\text{-}i) \qquad (6\text{-}14)............$$

حيث ترمز $\rho i\ (x\text{-}\iota)$ إلى معامل ارتباط درجة الفقرة بالدرجة الكلية بحذف الفقرة المعينة بمعامل الارتباط، $\sigma_i\ \sigma_x$ إلى الانحرافات المعيارية للدرجة الكلية ولدرجة الفقرة على التوالي.

معامل ارتباط بايسيريال:

عندما نرغب بان يكون توزيع المتغير الذي تقيسه الفقرة توزيعاً اعتدالياً، فمن الممكن اشتقاق صيغة للارتباط بين هذا المتغير ومحك توزيعه متصل مثل درجة الاختبار. وقد اشتق هذا الاحصائي في البداية من قبل بيرسون (Pearson , 1909) وأطلق عليه اسم معامل ارتباط بايسيريال، ويحسب باستخدام الصيغة:

$$(y / P)\ \frac{\mu_x - \mu_+}{\sigma_x} = \rho_{bis} \qquad (7\text{-}14)............$$

حيث ترمز μ_+ إلى متوسط درجة المحك للذين أجابو على الفقرة إجابة صحيحة، و μ_x هي متوسط درجة المحك للمفحوصين جميعهم، و σ_x إلى الإنحراف المعياري، وP إلى نسبة الذين أجابوا إجابة صحيحة على الفقرة، و Y إلى الإحداثي Y للمنحنى الاعتدالي المعياري عند الدرجة الزائية المرتبطة بقيمة p للفقرة. فعلى سبيل المثال للفقرة 1 في جدول 14–2 كانت $p_1 = 0.60$، وبما أن P_1 أكبر من0.50، وبالعودة إلى جدول المنحنى الاعتدالي المعياري في الفهرس A فإننا سنختار عمود الاحتمالات (أو المساحات إلى اليسار) للدرجات الزائية الموجبة. وقيمة المساحة الأقرب إلى 0.60 هي 0.599، والاحداثي المرتبط بها يساوي 0.3867، ومن بيانات جدول (14–2) يمكننا حساب قيم μ_+ , μ_x ، σ_x وباستخدام هذه القيم في المعادلة (14–7) نحصل على قيمة P_{bis} والمساوية لـ 0.36 و σ_x ويبين جدول (14–3) حساب قيم معاملات ارتباط بايسيريال للفقرات 1، 2، 3. لاحظ أن هذه القيم مختلفة قليلاً عن تلك الناتجة عن معامل ارتباط بوينت بايسيريال.

ورياضياً، فإن العلاقة بين معاملات ارتباط بايسيريال وبونيت بايسيريال هي:

$$\rho_{pbis} \sqrt{\frac{Pq}{Y}} = \rho_{bis} \quad (8\text{-}14)............$$

ولأن قيمة احداثيY على المنحنى الاعتدالي أقل من قيمة $\sqrt{Pq}$ ، فإن قيمة معامل بايسريال يكون على الأقل أكبر من ارتباط بونيت بايسيريال بـ $\frac{1}{5}$ للمتغيرات نفسها& Lord) (Novicck ,1968. ويكون هذا الفرق الكسري غير متحيز للفقرات متوسطة الصعوبة، وبانخفاض قيمة ρ إلى ما دون 0.25 أو ارتفاعها إلى أعلى من 75% فإن الفرق بين معاملي الارتباط يزداد بحدة. وقد أثبت ماغنسون (Magnusson) بيانياً أنه عند مدى الصعوبة المتطرف يكون معامل ارتباط بايسريال أكبر 4 مرات من معامل ارتباط بوينت بايسيريال بين درجة الفقرة والدرجة الكلية. لاحظ أن النسبة بين المعاملين في جدول (14-3) للفقرة 3 أكبر بـ (1.67) منه للفقرة 2 (1.23) أو الفقرة 1 (1.28). وظهر هذا بسبب انخفاض أثر تباين الفقرة الذي يعمل على تقييد قيمة معامل ارتباط بونيت بايسيريال. لذا فلمستخدمي الاختبارات الذين يقارنون بين تقارير تحليل الفقرات للدراسات المختلفة التي تستخدم صيغ معاملات ارتباط مختلفة عليهم تذكر أن معامل ارتباط بايسيريال أكبر وبانتظام من معامل ارتباط بونيت بايسيريال، وبالتالي فإن هنالك فروقات واضحة في قيم معالم تمييز الفقرة تعود إلى الصيغة المستخدمة في حساب معامل الارتباط لا إلى الفروقات النوعية بين الفقرات.

معامل ارتباط فاي:

عندما نريد حساب معامل ارتباط فقرات ثنائية التصحيح ودرجات محك ثنائي (مثل ناجح وراسب في برنامج إعادة التأهيل، أو خصائص ديموغرافية مثل الجنس) يمكننا استخدام معامل فاي المبين في الفصل الخامس. واستخدام آخر لهذا المعامل هو تحديد درجة الاستقرار في الاستجابات للفقرة نفسها ثنائية التصحيح للمفحوصين نفسهم في موقف آخر مختلف. ويكون استخدام معامل فاي مناسباً أكثر عندما تكون المتغيرات ثنائية حقيقية. فعند تشكيل مجموعات محكية باصطناع درجة قطع على توزيع المتصل، فإن هذا الإحصائي لا يسمح باستخدام تام للمعلومات المناسبة. بكلمات أخرى عند تصنيف المفحوصين الذين أعلى من درجة القطع بالدرجة 1 وأدنى منها بالدرجة صفر، فإن المعلومات الكمية حول الفروق بين الدرجات في النجاح والفشل ستضيع. ومحدد آخر محتمل لمعامل فاي هو أن قيمته تساوي p عندما تكون قيمة أ لكلا المتغيرين متساوية فقط (وذلك لأنه مثل معامل ارتباط بايسيريال النقطي يشتق مباشرة من معامل ارتباط بيرسون).

معامل ارتباط تتراشورك:

عندما يهتم مطوّر الاختبار بدرجة الارتباط بين متغيرين ثنائيين عندما يكون كلاهما ناجماً عن تحويل متغير متصل تحويل إلى متغير متقطع ثنائي. ومن المناسب في مثل هذه الحالات استخدام معامل ارتباط تتراشورك. وحساب هذا الاحصائي معقد جداً ونادراً ما يتم حسابه إلا في حالات الاعتقاد القوي به من قبل مطور الاختبار أكثر من أي معامل آخر مثل معامل فاي الذي لا يكون مناسباً للهدف. واحد هذه المواقف في حالة اجراء تحليل عاملي وتكوين مصفوفة الارتباطات الداخلية لفقرات ثنائية التصحيح. ولأن قيم معامل فاي مقيدة فيما عدا كون قيم P متساوية، فإن مثل هذه الارتباطات تكون أقل ملائمة من معاملات تتراشورك للتحليل العاملي. ولا تعرض صيغة معاملات تتراشورك بسهولة، ولكن يتوافر برامج محوسبة في العديد من الحقائب المحوسبة لحساب هذا المعامل انظر على سبيل المثال (Dixonietal ,1981).

مقارنة بين معاملات تمييز الفقرة:

تم عرض خمسة طرائق مختلفة في حساب تمييز الفقرة، ومن الواضح في بعض المواقف بسبب تصحيح المتغيرات، فإن إحدى التقنيات تكون أكثر ملائمة من غيرها. ومن المنطق طرح سؤال حول تشابه النتائج المحصلة من الطرائق المختلفة. وتناولت عدة دراسات تجريبية هذه القضية فقد استخدم انجل هارت (**Englehart , 1965**) استجابات 210 مفحوصين تقدموا لاختبار كفاءة المدرسة لعليا وباستخدام الدرجة الكلية كمحك، أجرى حساب معاملات تمييز متنوعة بما فيها D، فاي، بايسيريال ρ، بوينت بايسيريال ρ، وتتراشورك ρ. وعندما حسبت معاملات معاملات الارتباط بين هذه المعاملات تراوحت الارتباطات بين أزواج معاملات التمييز بين 0.85 إلى 0.99 ومن 0.90 إلى 0.99 على الصيغة الموازية ودونت دراسات مشابهة من قبل (Beuchert & Mendozaa , 1979, Oosterhof , 1976 , Findley , 1956) وفي معظم هذه الدراسات ظهرت الفجوات الأكبر للفقرات متطرفة الصعوبة.

كخلاصة، يمكن تقديم التوصيات المتعلقة باختيار طريقة تمييز الفقرة للفقرات ثنائية التصحيح الآتية:

1. عندما تكون الفقرات متوسطة الصعوبة يكون الفرق بين إحصائيات التمييز قليل، فإن كان المهم سهولة الحساب ينصح باستخدام D وإن كان الاهتمام في اختبار الدلالة الإحصائية ينصح باستخدام إحدى الطرائق الارتباطية.
2. وإن كان الهدف اختيار الفقرات متطرفة الصعوبة يوصى باستخدام معامل ارتباط بايسيريال فيما لو كان من المنطقي افتراض التوزيع الاعتدالي للسمة المؤثرة في الأداء.

3. عندما يشك مطور الاختبار باختلاف قدرة العينات المستقبلية عن عينة تحليل الفقرات، يوصى باستخدام معامل ارتباط بايسيريال إذ أن رتبة تمييز الفقرة في هذا الإحصائي يكون أكثر استقراراً من عينة لأخرى عندما تتغير القدرة. ويمكن القول بأسلوب آخر أن قيمة معامل ارتباط بايسيريال المنخفضة لعينة من أي مستوى قدرة يشير إلى أن الفقرة منخفضة في قوتها التمييزية، ولكن قيمة معامل بونيت بايسيريال المنخفضة للعينة منخفضة القدرة (أو عالية القدرة) قد تكون ببساطة دالة لصعوبة الفقرة وليست بالضرورة مؤشراً لضعف تمييزها.

4. وإن كان مطور الاختبار يثق ودون تحيز بأن العينات المستقبلية مشابهة في قدرتها لعينة التحليل، وكان الهدف اختيار فقرات ذات اتساق داخلي عالي فيفضل استخدام معامل ارتباط بونيت بايسيريال (Lord & Novick , 1968). ومع أن هذا لم يتم اثباته بشكل نهائي إلا أنه يبدو منطقياً من خلال حقيقة ان قيم معاملات بونيت بايسريال تكون أكبر للفقرات متوسطة الصعوبة، ومثل هذه الفقرات تسمح بأقصى تباين مشترك بين الفقرات وبالتالي أعلى قيمة لمعامل الفا.

5. وفي حالات التدريج الثنائي لكل من الفقرة والمحك، ينصح باستخدام معامل فاي أوتتراشورك، وحساب معامل فاي اسهل ولكنه محرر عندما تكون النسب غير متساوية لكلا المتغيرين. ويعتمد معامل تتراشورك على افتراض ان درجات الفقرة والمحك تظهر عند تحويل متغيرين توزيعها اعتدالي الى متغيرين ثنائيين. ونادراًماينصح باستخدام هذا المعامل كمقياس لتمييز الفقرة للتعقيدات في حسابه، ولكن يوصى به بقوة عندما يراد استخدام الارتباطات الداخلية في التحليل العاملي.

معاملات ثبات الفقرة وصدقها :

الفئة الثالثة من معالم الفقرة في دراسة تحليل الفقرة يعرف بمعامل ثبات الفقرة ومعامل صدقها. وكل منهما دالة لتباين درجة الفقرة وارتباط درجة الفقرة بالمحك. فان كان المحك المستخدم داخلي: الدرجة الكلية على الاختبار، فأن المعامل يعرف على انه $\rho_{ix}\,\sigma_i$ ، حيث تشير ρ_{ix} الى الارتباط بين درجة الفقرة والدرجة الكلية، ويطلق على هذا المعامل اسم معامل ثبات الفقرة، وتكتب هذه الصيغة بشكل أكثر شيوعاً للفقرات الثنائية $\rho_{ix}.\rho_i\sqrt{q_i}$ حيث ترمز ρ_{ix} إلى معامل ارتباط بونيت بايسيريال بين درجة الفقرة والدرجة الكلية، وعند استخدام محك خارجي فإن المعامل يحدد بـ $\rho_{iy}\,\sigma_i$، حيث ترمز ρ_{iy} إلى الارتباط بين درجة الفقرة ودرجة المحك الخارجي. وعندما يكون الهدف من اختيار الفقرات هو تحسين ثبات

درجات الاختبار أو صدقها يتم ذلك باختيار الفقرات التي تميز على المحك. وقد يقترح أحياناً استخدام معامل ثبات الفقرة (أو معامل صدق الفقرة) كبديل للارتباط البسيط بين درجات الفقرة والمحك، وذلك لأن تباين الفقرة يوازن بالفعل مساهمة فقرة معينة لثبات درجات الاختبار الكلي أو صدقه. وإن كان هنالك فقرتين (على سبيل المثال) لها ارتباطات متساوية مع الدرجة الكلية، وتباين أحداهما أكبر، فإن الفقرة ذات التباين الأكبر تساهم أكثر في ثبات درجات الاختبار لذلك يجب ملاحظة أنه كلما كانت الفقرات المختارة متوسطة في صعوبتها فإن هنالك ميزة إضافية عملية قليلة في استخدام معامل ثبات الفقرة بدلاً من معامل ارتباط درجة الفقرة بالدرجة الكلية.

ويكون معامل ثبات الفقرة أو صدقها مفيداً في بعض المواقف في بناء الاختبار. ويمكن تبين أنه يمكن التعبير عن تباين الدرجة الكلية بمجموع ثبات الفقرات، لذلك فإن:

$$\sigma_x = \sum(\rho_{ix}\,\sigma_i) \qquad (9\text{-}14)\ldots\ldots\ldots\ldots$$

وقد يفيد هذا في تحليل الفقرات عندما يضع مطور الاختبار أدنى قيمة مرغوبة لتباين الدرجة الكلية. وحالما تبدأ عملية اختبار الفقرات فإن مجموع معاملات ثبات الفقرات يزداد بإضافة كل فقرة حتى يظل الأدنى قيمة مرغوبة لتباين الدرجة الكلية، ويحسب هذا ببساطة أكبر بإضافة درجة الفقرة وإعادة حساب تباين الدرجة الخام الكلية عند إضافة كل فقرة.

بصورة مشابهة، فإن وضع مطور الاختبار أدنى قيمة لمعامل الاتساق الداخلي كما يقاس بمعامل الفا، فإنه يمكن إعادة حساب معامل الفا بعد إضافة كل فقرة باستخدام بيانات على مستوى الفقرة فقط من خلال الصيغة:

$$\rho_x = \frac{k}{k-1}\left(1-\frac{\sum\sigma_i^2}{\left(\sum\rho_{ix}\sigma_i\right)^2}\right) \qquad (10\text{-}14)\ldots\ldots\ldots\ldots$$

حيث ترمز K إلى عدد الفقرات المختارة.

وأخيراً ، أن كان مطور الاختبار يرغب في تخصيص أدنى قيمة لمعامل الصدق بين المجموعة الفرعية من الفقرات المختارة ومحك خارجي، فإن معامل الصدق بعد إضافة كل فقرة يحسب من البيانات على مستوى الفقرة على النحو التالي:

$$\rho_{xy} = \frac{\sum\rho_{xy}\,\sigma_i}{\sum\rho_{ix}\,\sigma_i} \qquad (11\text{-}14)\ldots\ldots\ldots\ldots$$

ويجب ملاحظة أن قيم التباين ومعامل الفا ومعامل الصدق المحسوبة من معاملات ثبات الفقرة تكون تقريبية، وذلك عندما نحصل على ρ_{ix} من تحليل الفقرة للفقرات جميعها من التجريب الميداني للفقرات لا من خلال الفقرات المختارة من المجموع الكلي. ولاشتقاق هذه صيغ العلاقة بين معالم الفقرة ومعالم الدرجة الكلية يمكن الرجوع إليها في (Gulliksen 1950) و (Lord & Novick ,1968).

إجراء دراسة تحليل الفقرات:

التجريب الميداني وتحليل الفقرات: نظرة عامة

لغاية الآن من الواضح أن خصائص الدرجة الكلية دالة للفقرات التي تؤلف الاختبار. لذا فإن أراد مطور الاختبار الحصول على درجات بأقل قدر من خطأ القياس، أو أنها ذات علاقة قوية بمحك الأداء أو قياسات لأبنية أخرى فلن يكون كافياً كتابة الفقرات ثم " يأمل بالأفضل " فحالما يتم كتابة الفقرات جميعها ومراجعتها والتجريب الأولي لها، ثم بعدها يمكن تطبيقها لاختبارها ميدانياً على عينة ملائمة من المفحوصين. ويحدد مطور الاختبار خلال تحليل الفقرات التي تؤدي وظيفتها وتلك غير الملائمة، وفي معظم الحالات يحتفظ مطور الاختبار بالفقرات الأولى، ويراجع الفقرات غيرالملائمة أو يقوم بحذفها. وفي تحليل الاختبار النموذجي على مطور الاختبار إجراء الخطوات الآتية:

1. حدد خصائص درجات الاختبار الأكثر أهمية.
2. حدد معالم الفقرة الأكثر ملائمة لهذه الخصائص.
3. طبق الفقرات على عينة مفحوصين ممثلة لمجتمع المفحوصين المستهدف في الاختبار.
4. احسب المعالم المذكورة في الخطوة الثانية لكل فقرة.
5. أسس خطة لاختيار الفقرات (أو تحديد الفقرات التي تحوي قصور ثم راجعها).
6. اختر المجموعة النهائية من الفقرات من الملف الإجمالي.
7. قيّم للتحقق من النتائج وذلك بإجراء دراسة صدق تقاطعي.

وفي الأجزاء السابقة من هذا الفصل حددنا معالم الفقرة التي تهمنا على الأغلب وعرفناها، وذلك لأهمية علاقتها بمعالم الدرجة الكلية، وكذلك عرضت الصيغ الرياضية اللازمة لحساب هذه المعالم. وفي الأجزاء الآتية سنركز على الخطوات 3، 5، 6، 7 بالتسلسل السابق نفسه.

حجم العينة:

لا توجد قاعدة مطلقة لأقل عدد من المفحوصين يلزم في تحليل الفقرات. ومن المؤكد أن تحليل اختبار يستخدم بكل واسع مثل GRE أو اختبارات الاستعداد أو التحصيل لمنشورة تجارياً يجب أن تعتمد على عينة ممثلة وكبيرة، وبالطبع آلاف من المفحوصين. بالمقابل طالب الدكتوراه الذي يطور أداة لأطروحته يجب أن يعتمد على عينة أصغر من هذه بكثير وكقاعدة عامة يمكن حساب معظم المعالم الموصوفة في هذا الفصل بعينات مستقرة نسبياً ومؤلفة من 200 مفحوص، ويعد هذا أقل عدد مرغوب. وقانون آخر أثبت علمياً هو أن يكون عدد المفحوصين من 5 إلى 10 أضعاف عدد الأفراد (Nunnaly ,1967). وإن استخدام مطوّر الاختبار هذا المؤشر فإنه سيستخدم على الأقل 100 مفحوص لكل فقرة. وحجوم العينات اللازم لحساب معالم الفقرة في نظرية IRT يتراوح من (200 إلى 1000) مفحوص بالاعتماد على النموذج المختار في التحليل. ومن الضروري لمطّور الاختبار ان يعرف طرائق الاختبار التي سيطبقها عندما يخطط للتطبيق الميداني، لذا فإنه سيتم التطبيق على عدد كاف من المفحوصين. نقطة أخرى يجب اخذها بعين الاعتبار هو الحاجة الى دراسة صدق تقاطعي، والذي سيتم عرضه باختصار، وتتطلب المرحلة النهائية في دراسة تحليل الفقرات اختبار افراد اخرين، بالاضافة الى الذين استخدمت استجاباتهم في حساب معالم الفقرة .

تأسيس خطة لاختبار الفقرات:

تاريخياً، هنالك قضية خلافية في تطوير الاختبارحول ماإذا كان اختيار الفقرات المرتبطة بالدرجة الكلية (والتي ينتج عنها درجات اختبار لها اتساق داخلي عالي) هو الافضل ام اختيارالفقرات الاقوى ارتباطاً بمحك خارجي. فمن جهة إن كانت الفقرات مؤلفة من عناصر مختلفة وتم اختيارها لان ارتباطهااعلى بمحك الخارجي، ولكن ارتباطاتها مع بعضها البعض قليل. وهنا يكون معنى وتفسير درجات الاختيار على انها مقياس للبناء معرضا للمساءلة (انظر Travers,1951) من جهة اخرى كلما كان ارتباط الفقرات مع بعضها البعض اكبر كلما كان التباين القليل المضاف الى المحك يعزى الى تجمع الفقرات. وعلى مطور الاختبار تذكر نقطة مفتاحية في تعيين المحك الذي سيستخدمه في تمييز الفقرات وهو أنه يجب اختبار المحك من خلال الهدف المحدد للاختبار والمشابه للمنظور الذي سيستخدمه. فكلما كان الهدف فريد لتنبؤ الاختبار بالاداء على محك مفرد كلما ازدادت منطقية اختيار الفقرات التي ترتبط بهذا المحك. مع ان هذه التفردية للهدف في استخدام الاختبار نادراً نسبياً، والاكثر شيوعاً هو تطوير الاختبار ليمثل نطاق سلوك أو التنبؤ بالاداء المستقبلي الذي لا يمكن تحديده بوساطة محك مفرد. وفي مثل هذه الحالات فان اختيار الفقرات التي تعظم القدرة التنبؤية للاختبار على محك معين في موقف تطبيقي محلي، ودون الحاجة الى ادلة لعلاقة الفقرات بالبناء الواسع الذي نهتم به، وهذا يؤدي الى محدودية فائدة الاختبار في مواقف اخرى حيث يختلف المحك او مجتمع المفحوصين قليلاً. مع ذلك فان العديد من مطوري الاختبارات يكتبون فقرات تقيس السمة المقصودة ، واختيار الفقرات التي من خلال تحليلها بالاعتماد على الدرجة الكلية

على الاختيار، ومن بعدها تقييم كيفية ارتباط درجات الاختبار كمقياس للبناء بمحك خارجي أكثر.

وفي اختيار الفقرات المجربة ميدانيا من اجل الاحتفاظ بها في الصورة النهائية للاختبار، وهنا يواجه مطور الاختبار بأحد المواقف الآتية :في البداية يوجد عدد فقرات اكبر بكثير من التي ستطبق خلال التطبيق الروتيني للاختبار، ولتوفير الوقت في العملية الاختبارية والتصحيح فان مستخدمي الاختبارات يريدون تطبيق أقل عدد ضروري من الفقرات . لذلك فإن المهمة تكمن في اختيار مجموعة فرعية من الفقرات تؤدي الى اكبر مساهمة في المستوى المرغوب لثبات الدرجات أوصدقها وبأقل عدد ممكن من الفقرات. وسيتم مناقشة اهمية معامل ثبات الفقرة (اومعامل صدقها) في الحال. وفي الموقف الثاني عندما يكون الملف الاختباري ليس كبير، ويريد مطور الاختبار الاحتفاظ بكل فقرة تساهم ايجابيا في معلم درجة الاختبار الذي يهتم به بشكل اساسي. وفي هذه الحالة يجب تأسيس ادنى حد لمعلم تمييز الفقرة ومن ثم الاحتفاظ بكل فقرة لها معلم تمييز اكبر من هذا الحد الادنى. واقترح ايبل (EBEL, 1965) محك يمكن استخدامه لتقييم معامل التمييز. ففي حالة المعاملات الارتباطية يختار مطور الاختبار كل فقرة لها ارتباط مع المحك اكبر من الصفر وبدلالة احصائية وفي حالة معاملات ارتباط فاي وبوينت بايسيريال. يمكن استخدام تقريب مناسب للخطأ المعياري لمعامل ارتباط بيرسون لتأسيس هذا المستوى، وذلك بحساب:

$$\frac{1}{\sqrt{1 - N}} = \hat{\sigma}_p \qquad (12\text{-}14)\ldots\ldots\ldots\ldots$$

حيث ترمز N الى حجم العينة، (وعند استخدام هذه الصيغة يكون حجم العينة على الاقل50). وغالباً ماتقع القيم الحرجة على بعد 2 انحراف معياري أعلى من الصفر، لذا فإنه لعينة حجمها 101 فإن

$$0.1 = \frac{1}{10} = \hat{\sigma}_p$$

و $0.00 + 2\ (0.1) = 0.20$، لذا فإننا سنحتفظ بكل فقرة لها قيمة معامل بوينت بايسيريال =0.20 أو أكبر.

والخطأ المعياري لمعامل بايسيريال يمكن حسابه من الصيغة:

$$\frac{\sqrt{(1\text{-}N) / Pq}}{Y} = \sigma_{bis} \qquad (13\text{-}14)\ldots\ldots\ldots\ldots$$

حيث يرمز p الى نسبة الذين اجابوا اجابة صحيحة على الفقرة و $q = (p-1)$، و N الى

حجم العينة و y الى احداثي المنحنى الاعتدالي عند قيمة p (Kurtz & mayo,1979) وثانية يمكن استخدام الخطأ المعياري هذا في إيجاد الحد الأعلى للفترة حول الصفر. لذا فإنه يمكن تحديد الفقرات التي ارتباطها أكبر من المتوقع عن طريق الصدفة. ويكون الخطأ المعياري لمعامل بايسيريال اقل عندما تكون p =0.50 ويزداد عندما تصبح قيم أ متطرفة. لذا فإن استقرار هذا الاحصائي من عينة لاخرى يتأثربقوة بصعوبة الفقرة.

وفي تطوير استراتيجية اختيار الفقرات فإن مطور الاختبار المبتدىء قد يتفاجأ من كيفية تأثير بيانات صعوبة الفقرة على القرار. وهناك عدة ملاحظات مناسبة لهذه القضية.

الأولى: تعد صعوبة الفقرة محك اساسي في اختيار الفقرات، ولكنه أقل أهمية من تمييز الفقرة في الاختبارات المعيارية .

الثانية: تتغير قيم p من عينة لاخرى. (يمكن حساب خطا المعاينة لصعوبة الفقرة باستخدام الخطأ المعياري للنسبة $\hat{\sigma}_p = \sqrt{N/Pq}$ ، حيث N هي حجم العينة .

الاخيرة: في الاختبارات المتوقع انها تميز بثبات على مدى واسع من القدرة يجب ان تكون فقراتها متوسطة الصعوبة وموحدة. وعلى مدى سنوات عديدة بنى ناشروا الاختبارات التجارية اختبارات تألفت تقريباً من فقرات ذات صعوبة منخفضة ومتوسطة وعالية وبالتدريج حل محل هذا التطبيق استراتيجية اختيار الفقرات من المدى متوسط الصعوبة مع التصحيح للتخمين لفقرات الاختيار من متعدد. وهناك دراستين تأثرتا بهذا التطبيق اجراها كل من لورد **(Lord 1952)**و كرونباخ وواريينجتون **(Cronbach &Warrington 1952)**،عموماً بينت الدراستين انه عندما يكون ارتباط الفقرات مع الدرجة الكلية متوسطاً فإن الفقرات متوسطة الصعوبة الموحدة تسمح بتمييز بثبات اكبر بين المفحوصين عبر مستويات القدرة جميعها على وجه التقريب اكثر مما هو في حالة تجمع فقرات عبر مدى اوسع من الصعوبة، واكثر تحديداً اقترح هنريسن **(Henryssen . 1971)** انه عندما يكون متوسط معامل بايسيريال بين الفقرة والدرجة الكلية في المدى بين 0.30 إلى 0.40 يجب ان يتراوح المستوى المثالي لصعوبة الفقرة بين 0.40 إلى 0.60، ولكن بازدياد معدل معامل بايسيريال اكثر من 0.60 يكون مدى اكبر من الصعوبة مقبولاً. ويمكن تمييز احد الاستثناءات لهذه الاستراتيجية، وذلك عند استخدام درجات الاختبار على وجه الحصر في اتخاذ قرارات حول المفحوصين عند النهاية العليا أو الدنيا للتوزيع، على سبيل المثال عندما يستخدم الاختبار في اختيار المتقدمين لبرنامج تدريبي متخصص وبالتنافس، فانه سيتم اختيار المتقدمين الاكثر تأهيلاً، وهنا يكون من المناسب اختيار الفقرات ذات قيمة أ المنخفضة (أي الفقرات الصعبة) لان هذه الفقرات تؤدي الى نتائج

تميز بين المفحوصين على القدرة المعينة، وعندما يتم بناء اختبار لاختيار فئة من المجتمع فإنه من المهم لمطوّر الاختبار ان يبين الاستخدام المقصود بوضوح، إذ ان فائدة الاختبار لمفحوصين عند مستوى قدرة اخر يكون أقل أهمية.

استخدام بيانات تحليل الفقرة في مراجعة الاختبار:

يستهدف عرض دراسة الحالة في تحليل الفقرة التالية لتوضح

1/ اكتشاف الفقرات التي تتضمن خلل او قصور.

2/ استخدام بيانات استجابات المفحوصين في تشخيص القصور المحتمل في الفقرات، وتبين بيانات الدراسة استجابات 50 مفحوص على 35 فقرة في اختبار صفي لمرحلة الدراسة الأولية وباستخدام تدريج معياري المرجع واستخدم المعلم محك داخلي (درجة الاختبار الكلية) في حساب معاملات تمييز الفقرة، ويبين هذا المثال بيانات الفقرات من 21 إلى 35 .

ويمثل جدول (14-4) مخرجات برنامج احصائي يجري تحليل للفقرات تبين المخرجات اثنين من احصائيات تمييز الفقرة، وصعوبة الفقرة ونسبة المفحوصين الذين اختاروا كل بديل اجابة.

ويفيد تمييز الفقرة في هذا الموقف في تحديد الفقرة " المشكلة "، ويفيد مستوى الصعوبة ونمط توزيع الاستجابة في تشخيص القصور في بناء الفقرات التي ينتج عنها تمييز ضعيف وبتطبيق محك ايبل على قيم D فإننا نعد الفقرات 22، 23، 28، 29، 32، 34، 35، فقرات جيدة لان قيم D لها تتجاوز 0.40، وعند تحديد قيمة معامل بوينت بايسيريال عند .... + $\acute{\sigma}_p$ فإن اقل قيمة مقبولة لمعامل بوينت بايسيريال لهذه العينة =0.29 (لعينة 50 مفحوص)، لذا فإننا سنختار الفقرات نفسها التي اخترناها بناءً على قيم D بالاضافة الى الفقرة 30 وتشير هذه النجمة بجانب الفقرة في الجدول (14–4) إلى ان الفقرة مقبولة على كلاً من محكي التمييز، ومثل هذه الفقرات يتم الاحتفاظ بها دون مراجعتها. ولتحديد الفقرات التي بحاجة الى مراجعة سنستخدم محك ايبل من خلال قيمة D اذ تكون اقل من0.20 ، وحددت هذه الفقرات بوضع خط تحتها في جدول 14–4، وفي اختبارنا كانت الفقرات المشكلة ستة وهي: 21، 24، 25، 26، 31، 33. (لاحظ ان الفقرة 30 هي الفقرة الوحيدة التي لها قيمة معامل بوينت بايسيريال مقبولة، ولكن قيمة D غير مقبولة. وفي هذه الحالة نكون أكثر ميلاً للاسترشاد بمعامل بوينت بايسيريال الذي يعتمد على استجابات المفحوصين جميعهم.

استجابات الفقرة %						الصعوبة p	معامل التمييز	معامل ارتباط بوينت بايسيريال
الفقرة	1	2	3	4	الحذف			
21	24	4	52	16+	4	0.16	0.00	0.06-
*22	4	40	56+	0	0	0.56	0.67	0.48
*23	0	76+	12	12	0	0.76	0.50	0.45
24	4	28+	28	32	8	0.28	0.17-	-0.12
25	16	12	0	72+	0	0.72	0.17-	-0.29
26	0	4	52	44+	0	0.44	0.00	-0.11
27	92+	0	8	0	0	0.92	0.33	0.45
*28	8	68+	0	20	4	0.68	0.83	0.61
*29	24	12	56+	8	0	0.56	0.50	0.46
30	88+	0	0	8	4	0.88	0.17	0.31
31	68+	12	4	16	0	0.68	0.17	0.15
*32	20	20	8	52+	0	0.52	1.00	0.73
33	8	16	60+	16	0	0.60	0.00	0.06
*34	20	20	8	52+	0	0.52	0.83	0.59
*35	80+	0	0	4	16	0.80	0.50	0.43

جدول (14 – 4) توضيح لنتائج تحليل الفقرة لمفحوصين عددهم 50
على الفقرات من 21 إلى 50 في اختبار عدد فقراته 50 3

+ الاجابة الصحيحة   × فقرة تميز بفعالية   – فقرة تمييزها ضعيف

والفقرة 21 لها تمييز سالب وفقاً لمعامل بوينت بايسيريال وتظهر هذه الفقرة شاذة جداً بين الفقرات جميعها إذ أن مستوى صعوبتها (P =0.16) وعند فحص توزيع الاستجابات نرى أن 52% من المفحوصين إختاروا الاستجابة 3 بدلاً من 4 والتي هي الاجابة الصحيحة، ولان هذه نسبة عالية وغير طبيعية في اختيار الاستجابة الخاطئة نفسها، فهنالك احتمال منطقي لان تكون الفقرة خاطئة (miskeyed) وفحص محتوى الفقرة أظهر أن هذه الحالة غريبة.

كذلك كان تمييز الفقرة 24 سالباً ويظهر انها اكثر صعوبة مما هو متوقع (P = 0.28) ولكن تتوزع استجابات المفحوصين بالتساوي عبر البدائل الثلاث ويظهر أنها كانت عشوائية، ويشير العدد الاكبر من المحذوفات الى ان المفحوصين مرتبكين حول ما يسألون عنه ومثل هذا النمط من الاستجابات يقترح احتمالات ثلاث لها: الكلمات غامضة في متن الفقرة لذا فإن المفحوصين لم يفهموا السؤال، او ان الفقرة تغطي محتوى غير مألوف لهؤلاء المفحوصين، وأنه لا يوجد اجابة صحيحة في هذه الفقرة أو اعادة صياغتها بشكل كامل.

ويبدو أن الفقرات 25، 26 تتضمن مشاكل في المحتوى او تركيب أحد بدائل الاجابة. فالفقرة 25 الاضعف تمييزاً لها مستوى صعوبة (P =0.72) أعلى قليلاً من المثالي. فمن توزيع الاستجابة نرى ان البديل 3 لم يؤشر عليه واحد من المفحوصين. والاستنتاج من الاخطاء من الواضح أنه يزيد فرصة اختيار الاجابة الصحيحة من قبل المفحوصين الاقل قدرة عن طريق التخمين، وفي مراجعة الفقرة نجد أنه يجب استبدال البديل 3 ببديل اكثر منطقية على أمل أن يجذب المفحوصين غير المتأكدين من الاجابة الصحيحة. وينصح ايضاً مراجعة نموذج إجابة المفحوصين ذوو الدرجات العالية وأخفقوا في الاجابة الصحيحة على هذه الفقرة. لتحديد ما إذا وجدوا بديلاً معيناً جذاباً وتحتوي الفقرة 26 خطائين غير وظيفيين (البديلين 1 ، 2)، والمشكلة الاكبر في هذه الفقرة هو أن 52% من المفحوصين اختاروا الاجابة الخاطئة نفسها. وواضح أنه يجب مراجعة محتوى البديل الصحيح للتأكد من دقته،اضافة الى انه يجب مراجعة البديل 3 كي يصبح أقل جاذبية، ويجب اعدة كتابة البديلين 1،2 لتصبح أكثر جاذبية. ويجب ملاحظة أن الغموض في متن الفقرة أحياناً يجعل نسبة كبيرة من الفمحوصين تنجذب الى استجابة خاطئة معينة، لذا يجب التحقق من هذه الاحتمالية.

ولم تقدم نتائج تحليل الفقرات 31، 33 مؤشراً معيناً لضعف تمييزها. لذا فإن التمحيص الجيد لمحتوى هذه الفقرات والاستجابات الخاطئة التي اختارتها الفئة العليا قد تكون ضرورية لتحديد المشكلة.

اخيراً، حتى الفقرة 35 التي يظهر ان لها تمييز فعال عالي، فإن هذا تم تحقيقه لان نسبة اساسية من المفحوصين لم يجيبوا على هذه الفقرة لانها الفقرة الاخيرة في الاختبار. لذا نتوقع تمديد الفترة الاختبارية قليلاً لاعطاء الوقت الكافي للمفحوصين كافة ليكملوا الاختبار.

وفي العملية الاختبارية الصفية فإن معظم الجهات المسؤولة توصي باستخدام بيانات تحليل الاختبار كاساس في مراجعة الاختبار مستقبلاً، ولكنهم لا ينصحون بحذف الفقرات

عند حساب درجات الصف الحالي الذي اختبر بها، فالطلبة عند تقدمهم لاختبار فإنهم يفترضون ان الفقرات جميعها تحسب في الدرجة، وهم لا يعتبرون التصحيح عادلاً فيما لو اختار المعلم مجموعة فرعية من الفقرات لاحتساب درجتها في الاختبار. وقد وضح كوكس (COX , 1965) ان بناء الاختبار على اساس معاملات تمييزها فقط لا تكون مقياساً صادقاً للاهداف التعليمية الموجودة في الخطة الدراسية. وفي تطوير صيغة تجريبية لاختبار لاستخدامه في التقويم او البحوث فإن درجات المفحوصين الذين اشتركوا في دراسة تحليل الفقرات غير مهمين بحد ذاتهم، ويمكن لمطور الاختبار في هذه الحالة حذف الفقرات التي تحوي خلل او قصور، لذلك فإنه ينصح بانتاج عدد أكبر من الفقرات اكبر مما نحتاج لكل هدف أو مجال محتوى وتجريبها ميدانياً للتأكد من توافر عدد كافي وملائم من الفقرات الجيدة للصيغة النهائية للاختبار والمتوازنة بشكل مناسب للمحتوى. واخيراً ان كان المعلم مهتماً بحفظ الاسئلة في مصرف اسئلة فانه يمكن استخدام بيانات التحليل لتحديد الفقرات الجاهزة للاستخدام الفوري، بالاضافة الى استخدام هذه البيانات في مراجعة الفقرات الخاطئة التي يمكن التحقق منها بتجربة ميدانية ثانية. وقد بيّن كل من لانج وليهمان وميهرنز (Lange, lehmann , Mehrens, 1965) ان اعداد فقرات جديدة يتطلب خمسة اضعاف الوقت اللازم لمراجعة الفقرات المتوافرة، لذا فإن اهمال الفقرات التي تحوي قصور دون محاولة مراجعتها يعد اجراء غير فعّال فيما لو احتاج باني الاختبار صيغ اختبارية تغطي المحتوى نفسه مستقبلاً.

دراسة الصدق التقاطعي :

عندما تختار الفقرات بناء على محك احصائي باستخدام استجابات عينة معينة، يجب ان يكون الاختبار اكثر فعالية لهذه العينة اكثر مما هو لأي عينة اخرى من المفحوصين، وتم برهنة هذا من قبل كيورتون (Curetom , 1950) الذي وصف تطوير اختبار اختار منه الفقرات التي كان اداء الفئة ذات متوسط علامات GPA أعلى من اداء الفئة منخفضة متوسط الدرجات GPA ثم استخدم الاستجابات على الفقرات نفسها وأعاد حساب الدرجة الكلية للمفحوصين جميعهم على الفقرات المختارة حسب معامل ارتباطها مع GPA، وكان معامل الارتباط ‎-0.80، وبعد ذلك بفترة أدرك كيورتون ان فقرات الاختبار كانت على بطاقة في وعاء متحرك، وكانت استجابات المفحوصين من خلال اسقاط بطاقات على طاولة ومكافئة الطالب في حالة ظهور الوجه للاعلى في البطاقة، وكانت فرصة ظهور الوجه لاعلى أو اسفل متساوية في كل رمية. وبالاختيار العشوائي كان وجه البطاقات لأعلى في الفئة عالية GPA اكثر مما هو

للفئة المتدنية، وهذه الفقرات هي التي احتفظ بها نتيجة تحليل الفقرات وعندما حسبت درجات هذه الفقرات فقط فإن درجات المفحوصين الكلية حسبت من البيانات نفسها ظهر ان الاداء كان اكثر ارتباطاً بمتوسط GPA. ومن الواضح ان مثل هذه العلاقة قد تختفي فيما لو أعيدت الدراسة عندما يتم رمي البطاقات ثانية، كذلك فإن الصدق التقاطعي مشابه لوضع الفقرات في محرك ورميها ثانية لرؤية ما إذا أدت الفقرات وظيفتها ثانية بفعالية. ولاجراء دراسة صدق تقاطعي يستخدم مطور الاختبار الفقرات التي اختيرت من نتائج تحليل الفقرات فقط، وتطبق هذه الفقرات على مجموعة ثانية مستقلة من المفحوصين وثبات و / أو صدق هذه الدرجات تحدد باستخدام الطرائق الموصوفة في الفصلين السابع والعاشر على التوالي.

وبسبب الجهد المبذول في تطبيق الاختبار فمن الشائع في دراسة تحليل الفقرات/ الصدق التقاطعي جمع بيانات كلا الحالتين في جلسة اختبارية واحدة، ويتم هذا بتطبيق الفقرات جميعها في الملف الاختباري على المفحوصين جميعهم. ويؤشر الى كل ورقة اختبارية عشوائياً إما الى تحليل الفقرات أو الى دراسة الصدق التقاطعي، فعلى سبيل المثال لو اختبر 400 مفحوصاً في30 فقرة، تقسم اوراق الاجابة الى مجموعتين كل منهما تتألف من 200 ورقة وبعدها يستخدم مطور الاختبار احدى المجموعتين في تحليل الفقرات، وعلى اساس هذا التحليل افترض ان 20 فقرة اختيرت في الصيغة النهائية للاختبار وفي المجموعة الثانية المؤلفة من 200 ورقة تستخدم 20 فقرة فقط في دراسة الصدق التقاطعي في حساب الثبات والصدق. وفي بعض الاحيان قد يرغب مطور الاختبار دراسة ما إذا تم الحصول على نتائج مشابهة بغض النظر عن اي المجموعتين استخدمت في تحليل الفقرات أو الصدق التقاطعي، ويمكن اجراء هذا باستخدام العينة 1 في تحليل الفقرات والعينة 2 في دراسة الصدق ثم اعادة الدراسة باستخدام العينة 2 في تحليل الفقرات وعينة 1 في دراسة الصدق التقاطعي، وهذا ما يعرف بـ الصدق التقاطعي المضاعف

ومن الواضح ان الحاجة للصدق التقاطعي يعني ان عدد أكبر من المفحوصين يجب توافره عما يلزم لتحليل الفقرات. احد الاسئلة التي تظهر احياناً في تحليل الفقرة / الصدق التقاطعي يتعلق بنسبة المفحوصين اللازم لكلا العينتين، مع ان التجزئة 50:50 المستخدمة في المثال السابق شائعة، ففي حالات اخرى تكون التقسيمات اكثر منطقية، فإن كان الملف الاختباري مثلاً مؤلف من 50 فقرة وتوافر 400 مفحوص فقط فإنه يفضل استخدام نسبة اكبر من المجموعة في تحليل الفقرات ونسبة أقل في الصدق التقاطعي وهنا يستخدم على الاقل 250 مفحوص في تحليل الفقرات و 150 في الصدق التقاطعي (لوحظ ان هذه التجزئة حددت بوساطة القاعدة: 5 مفحوصين على الاقل لكل فقرة في تحليل الفقرات)

تحليل الفقرات في الاختبارات المحكية

الهدف الشائع للاختبارات المحكية هو تقييم الاداء على مجموعة مهام ممثلة لمجال محدد تماماً، لذلك فان مطوروا الاختبارات المحكية يستخدمون بثبات بعض التقنيات لدعم صدق المحتوى للفقرات من خلال احكام الخبراء التي نوقشت في الفصل 10، لذا فإن فحص بيانات الاستجابات التجريبية تكون مفيدة في تطوير الاختبارات المحكية وفي فحص بيانات الاجابة على الاختبار فإن مطور الاختبار نادراً ما ينظر الى الفقرات التي تتضمن قصور فقط وبدلاً من ذلك فإن العملية التدريسية الكلية، وخطة تطوير الاختبار، والفقرات تكون خاضعة للتدقيق، فعندما لا يكون اداء المفحوصين كما هو متوقع على فقرة معينة فان الفشل قد يكون لعدم ملائمة التدريس او خصائص الاختبار أو الاهداف او تركيب الفقرة. وهدف تحليل الفقرات في الاختبارات محكية المرجع هو التحقق من العوامل الدخيلة لنطاق معين تعزى الى الاداء على فقرات الاختبار، لهذا السبب فان اوائل المنادين بالاختبارات المحكية اشاروا الى ان احصائيات الفقرة لا يجب ان تكون دالة لتباين درجة الفقرة لمجموعة مفحوصين، واحدة انظر: (Millman & popham ,1974 , popham & Husek 1969) لذا فإن معاملات تمييز الفقرة الموصوفة في الاجزاء السابقة من الفصل لا تناسب الاختبارات المحكية كما هو الحال للاختبارات المعيارية، فانه يجب اجراء تحليل الفقرات للاختبارات المحكية بعد تحديد الاهداف بوضوح وعلى وجه الخصوص يجب ان يعرف مطور الاختبار لماذا يحتاج الى معلومات عن استجابات الفقرة وكيفية استخدامها، وفي مواقف معينة فان واحداً أو اكثر من هذه الاسئلة تكون مناسبة.

1/ ما مستوى صعوبة الفقرات؟

2/ هل الفقرة حساسة للتدريس؟ مثل هل تميز بين الذي درس والذي لم يدرس؟

3/ هل هنالك اتفاق عبر انماط الاستجابة لفقرات معينة، كما هو مفترض من مواصفات الاختبار؟

اقترح العديد من الاحصائيات للاجابة عن هذه الاسئلة. واحد أو أكثر من هذه الطرائق معروض هنا. وهذه الطرائق المعروضة اختيرت لعمومية تطبيقاتها (قابليتها للتطبيق على العموم وسهولة حسابها، وتفسيرها واضح غير غامض، والقارئ المهتم بتفاصيل اكثر لهذه الطرائق وغيرها من طرق تحليل فقرات الاختبارات المحكية عليه الرجوع الى

(Berk , 1980 b ,Harris , Pearlman & Wilcox1977 ,Harris, alkin, & popham 1974, & Lord ,1980)

صعوبة الفقرات:

يعرف مستوى صعوبة الفقرة في الاختبارات المحكية عموماً على أنها نسبة المفحوصين الذين أجابوا إجابة صحيحة على الفقرة (P).

والمناقشة السابقة لهذا المفهوم مناسبة عدا ان مطور الاختبار المحكي لا يهتم باختيار الفقرات التي تعظم التباين. مع ذلك فإن فحص صعوبة الفقرة يبقى مهماً. فمن المنطق تحديد متوسط أو وسيط مستوى صعوبة كل مجموعة من الفقرات التي تقيس هدف معين. وتفيد هذه القيمة في تقييم فاعلية التدريس للهدف او ملائمة المواصفات للفقرة، على سبيل المثال: الفقرات السهلة جداً في الاختبار القبلي تجعل مطور الاختبار يتساءل عما اذا كان التدريس لهذا المحتوى ضروري أو فيه اسهاب لهذه الفئة من المفحوصين، بينما الفقرات متطرفة الصعوبة لمجموعة بعد التدريس قد تشير الى ان التدريس كان غير فاعل أو ان المواصفات تتضمن محتوى او عملية غير مغطاة بالأهداف التعليمية والفقرة الواحدة السهلة جداً أو الصعبة جداً بالنسبة للفقرات الاخرى المتعلقة بالهدف نفسه يجب فحصها من حيث الخلل او القصور التقني او اي مؤشرات غير مقصودة او اخطاء مفتاحية، أو غموض في الكلمات التي قد تؤثر على الصعوبة بغض النظر عن المحتوى، وقد تقدم الفقرات متنوعة الصعوبة التي تقيس هدف معين معلومات مفيدة، فعندما تكون مستويات صعوبة هذه الفقرات متنوعة جداً (ليس بسبب أواثنين من أسباب القصور التقني) فإنه ينبغي مراجعة الاهداف أو مواصفات الفقرة على الترتيب .

حساسية التدريس:

يعد مقياس حساسية التدريس للفقــرة بالاساس مقياساً لجودة تمييز الفقرة بين المفحوصين الذين حصلـــوا على التدريس ومن لم يدرسوا. واقتــرح كوكس وفارجاس (Cox& vargas 1966) طريقة يتم فيها اختبار المجموعة قبل التدريس وبعده، ويحدد احصائي التمييز على النحو الآتي:

$$D = p_{\text{بعد}} - P_{\text{قبل}} \qquad \text{(4-14)}$$

حيث ترمز P بعد الى نسبة الذين اجابوا اجابة صحيحة على الفقرة في الاختبار البعدي وP قبل الى النسبة نفسها في الاختبار القبلي.

وتتراوح قيم D من −1 الى +1 مع تفضيل القيم الموجبة العالية. وصيغة اخرى تستخدم استجابات مجموعتين مستقلتين احداهما تعرضت للتدريس والاخرى لم تحصل عليه. واقترح بريئان (Brenan, 1972) طريقة تتطلب درجة قطع السيطرة :

$(n_2/L - n_1/U) = B$ ....(5-14)

حيث يرمز n_1 الى عدد الذين حصلوا على درجة أعلى من درجة القطع و n_2 أقل منها، و U الى عدد الذين درجاتهم أعلى من درجة القطع واجاباتهم على الفقرة صحيحة، وL إلى عدد الذين درجاتهم دون درجة القطع وإجابتهم صحيحة. ولان B هي الفرق بين النسبتين فإنها تتراوح بين ‎-0.1 و‎+ 1 مع تفضيل القيم الموجبة العالية، وملاحظة ان B مقياس لقدرة الفقرة على التمييز عند نقطة قطع معينة.

وعندما يريد مطور الاختبار تحديد اي الفقرات اكثرفاعلية نسبية او أقل في التمييز بين مجموعات الدارسين وغير الدارسين يمكن استخدام عدة طرائق إرتباطية. فقد اقترح بيرك **(Berk ,1980)** استخدام طريقة ارتباطية اشتقها ساوب **(Saupe ,1966)**. في اختيار الفقرات والأدوات المصممة لقياس التغير. ويتطلب استخدام هذه الصيغة تطبيق كل فقرة على المجموعة نفسها في تصميم اختبار قبلي- اختبار بعدي. ويحسب لكل مفحوص درجة التغير للفقرة وتكون (1، أو صفر، أو ‎-1) على التوالي، مشيراً الى كسب 1 نقطة، اولا كسب، او خسارة نقطة على التوالي. بالاضافة الى حساب درجات المفحوصين الكلية على الاختبارين القبلي والبعدي، وكذلك درجة التغير الكلي:

$D_{\text{الكلية}} = X - Y$ .... (16-14)

حيث ترمز y الدرجة الكلية على الاختبار البعدي، و x الى الدرجة الكلية على الاختبار القبلي، ولكل فقرة حساب معامل ارتباط بيرسون بين درجة التغير للفقرة ودرجة التغير الكلية. واقترح ساوب **(Saupe 1966)** صيغة حسابية بديلة لهذا الارتباط يؤدي الى نتائج مشابهة لمعامل ارتباط بيرسون بين درجة التغير على الفقرة ودرجة التغير الكلية .

ووصف ميلمان **(Millman ,1974)** طريقتان غير السابقة لتقييم حساسية التدريس يستخدم فيها مطوّر الاختبار بيانات استجابات الفقرات لمجموعتين منفصلتين احداهمامفحوصين تم تدريسهم والاخرى لم تدرس، ومن خلال معامل الارتباط الجزئي او الانحدار المتدرج فإن كلا الطريقتين تسمح لمطور الاختبار ليبدأ بمجموعة فرعية صغيرة في الفقرات وتقييم تحسين الاختبار لحساسية التدريس مع اضافة كل فقرة للمجموعة .

ويجب ملاحظة ان اختيار (أوحتى مراجعة) الفقرات على اساس معاملات حساسية التعليم قد تتضارب مع هدف الاختبار المحكي الذي طور في الاصل، فإن كان الملف الاصلي للفقرات يمثل أهداف حددت من قبل، وتتألف مما يجب ان يعرف المفحوص. فالفقرات

الحساسة للتدريس يجب ان تتكون من مجموعة فرعية لنطاق المحتوى (مثل ماذا دُرس بفاعلية) ، افترض على سبيل المثال أنه ولهدف واحد مهم لم يكن هناك تدريس ولم يظهر تعلم، فإذا كانت الفقرات المختارة على اساس قيم عالية لـ (D) (Cox & Vargas,1966) أو B (Brenan ,1972) فإن فقرات هذا الهدف جميعها يجب حذفها من الاختبار. ومن الواضح أن حذف هذه الفقرات لا يحسن صدق المحتوى للاختبار محكي المرجع، فضلاً عن ان طريقة مثل طريقة ساوب لتغير درجة الفقرة صممت بالفعل للقياس معياري المرجع، ومثل هذا المعامل ينتج عنه اختيار فقرات لها تباين في درجات التغير والذي ليس هدفاً اساسياً لمطور الاختبار المحكي المرجع لذا فليس من المنطق حساب معاملات حساسية التدريس روتيناً لكل اختبار محكي ومن المهم المناداة لاستخدام مثل هذه المعلومات والتمييز بأن اختيار الفقرات على أساس مثل هذه المحكات الاحصائية لا تتطابق مع أساليب السيطرة على النطاق في القياس. وقد تزودنا معاملات حساسية التدريس بمعلومات مفيدة حول نوع المحتوى الذي تم تدريسه بفعالية او بدون فعالية في برنامج تدريسي معين. واستخدام بيانات تحليل الفقرات بهذه الطريقة تفترض مسبقاً ان مستخدم الاختباريضع ثقته في نوعية الفقرات.

معاملات الموافقة :

في العديد من الحالات يكون اهتمام مطوّر الاختبار في دراسة تشابه استجابات مجموعة واحدة من المفحوصين لكل زوج من الفقرات كتبت بالخصائص نفسها. يبدو هذا منطقياً في المواقف التي يتم فيها تطوير اختبارات متتابعة من خلال تطوير الفقرات عشوائياً بوساطة مجموعة فقرات اختبرت ميدانياً. وهنا يريد مطور الاختبار معرفة ما إذا كان بالامكان افتراض التبادلية لهذه الفقرات، ويمكن في هذه الحالة ترتيب بيانات كل زوج من الاستجابات في جدول رباعي مثل جدول (14-5). وقد ناقش كل من هاريس وبيرلمان (Harris&perrIman 1977) معاملات احصائية متنوعة يمكن تطبيقها على مثل هذه البيانات، ويعتمد الاختبار الاحصائي المناسب على ماذا يريد مطور الاختبار ان يعرف فيجب ان يطرح مطور الاختبار بعض الاسئلة التوضيحية، ووصف كل من هاريس وبيرلمان عدد قليل من المعاملات التي يمكن استخدامها لتوضيح اسلوب تحليل الفقرات هذا. ووصف ويلكوكس (Wilcox) نماذج احتمالية اضافية لاختبار تكافؤ الفقرات، وهذه النماذج اكثر تعقيداً لانها تأخذ بعين الاعتبار احتمالية الاستجابات غير المناسبة من المفحوصين (وهو ان المفحوص قد يعرف الاجابة ولكنه قدم اجابة خاطئة أو العكس) والطرائق الموصوفة هنا لا تسمح بمثل هذه الامكانية.

وكمثال ابتدائي، افترض ان مطور الاختبار يريد معرفة " ما إذا كانت فقرتين تقيسان الشيء نفسه " وهنا من المناسب استخدام اختبار للاستقلال الاحصائي للاستجابات على الفقرتين ، ويمكن حساب احصائي مربع كاي بالصيغة:

$$\chi^2 = \frac{2(bc - ad)n}{(a+c)\ (b+d)\ (d+c)\ (b+a)} \qquad (17\text{-}14)$$

حيث ترمز: n الى عدد افراد العينة و d , c, b, a, هي تكرارات الخلايا في الجدول (14-5) وهذا الحساب موضح في جدول (14-6) واحصائي χ^2 المحسوب يقارن بقيمة χ^2 عند درجة حرية 1 ومستوى الفا المرغوب .

جدول (14-5): جدول رباعي الخلايا يمثل استجابات على ازواج الفقرات

الفقرة 2		الفقرة 1 +	الفقرة 1 −
	+	a	b
	−	c	d

(a)

الفقرة 2		الفقرة 1 +	الفقرة 1 −
	+	20	5
	−	5	20

(b) تكرارات استجابات 50 مفحوص

جدول (14-6): تطبيقات توضيحية لمعاملات الموافقة والارتباط لبيانات فقرات جدول (14 - 5 ب).

سؤال توضيحي	الصيغة	الحساب	التفسير التوضيحي
هل تقيس الفقرتين الشيء نفسه؟	$\chi^2=\frac{(bc-ad)^2 n}{(c+a)(d+b)(d+c)(a+b)}$	$\frac{(25-400)^2 50}{(30)(30)(30)(30)}$ =8.68	قيمة χ^2 المحسوبة تتجاوز الاختبار الاحصائي بـ 3.84 (لـ α=0.05) لذلك فان الاعتمادية لاستجابات هاتين الفقرتين دالة، وهذا يدعم ان الفقرتين تقيسان المهارة او المحتوى نفسه
ما نسبة المفحوصين الذين نجحوا أو فشلوا في كلا الفقرتين؟	P=(a+d)/η	50/(20+20) = 0.80	يوجد اتساق في الاداء على الفقرتين لـ 80 % من المفحوصين
هل أداء المفحوصين افضل وبدلالة على احدى الفقرات من الاخرى؟	$\chi^2=\frac{(\lvert c-b\rvert-1)^2}{c+b}$	$\frac{(\lvert 5-5\rvert-1)^2}{5+5}$ = 0.10	قيمة χ^2 المحسوبة لا تتجاوز الاختبار الاحصائي بـ (3.84 عند α= 0.05) وفشلنا في ايجاد الفرق الدال في صعوبات الفقرتين وبالتالي لا نستطيع استنتاج ان المفحوصين تعلموا المحتوى المقاس باحدى الفقرتين افضل من المحتوى المقاس بالفقرة الاخرى

فان كانت القيمة المحسوبة للاختبار الاحصائي دالة (كما وجد في مثالنا) فان الاستجابات على الفقرتين نقول انها معتمدة أو مرتبطة. والفشل بايجاد ارتباط دال بين الاستجابات على فقرتين مكتوبتين من خصائص الفقرات نفسها، تطرح سؤالاً حول ملائمة المواصفات نفسها أو التنقية النوعية لاحدى الفقرتين المستخدمتين او كلاهما

عند ايجاد تشابه في استجابات المفحوصين اكثر مما هو متوقع بالصدفة، فقد يرغب مطور الاختبار بعدها حساب احصائي يصف درجة التشابه. احد الاحصائيات والقابل للتفسير الفوري هو ببساطة نسبة الاتفاق n/ d + a، وهذه عبارة عن نسبة المفحوصين الذين اجاباتهم متسقة على الفقرتين انظر جدول (14-6) للحساب التوضيحي والتفسير لهذا

الاحصائي، وقد فضل كل من هاريس وبيرلمان هذا الاحصائي وذلك لانه يمكن ايجاد متوسطه وتفسيره معنوياًعبر أزواج فقرات عديدة ولانه تقدير غير متحيز. واحصائيات اخرى ملائمة لوصف درجة الاتفاق في الاستجابات لزوج الفقرات هو معامل كابا (وصف في فصل 8) ومعامل Q ليول الذي اوصى به هاريس وبيرلمان او معامل فاي، وسؤال مختلف قليلاً هو ما غذا كانت صعوبات فقرتين متساوية في مجتمع المفحوصين أعد صياغة السؤال بطريقة اخرى:

هل الفروق الملاحظة في الصعوبة صغيرة لدرجة يمكن ان تعزى الى خطأ المعاينة ومثل هذا السؤال يكون مناسباً فيما لو افترض مطور الاختبار ان البرنامج التعليمي ومحتوى الفقرتين تم تعليمها بالتساوي، ويرغب مطور الاختبار التحقق من هذا الغرض واقترح هاريس وبيرلمان لمثل هذه الحالة احصائي مربع كاي، وبالصيغة:

$$\chi^2 = \frac{^2(1\text{-IC-b}1)}{C+b} \qquad .... \qquad (8\text{-}14)$$

وثانية، يقارن هذا الا حصائي المحسوب بقيمة مربع كاي عند درجة حرية 1 ومستوى الفا المرغوب، وتشير القيمة الدالة لهذا الاحصائي الى ان الفرق في صعوبات الفقرتين لعينة المفحوصين اكبر مما يعزي للصدفة (انظر جدول 14 – 6) للحساب المطبق على الفقرتين في المثال) .

بشكل عام، قد تكون مناسبة عندما تظهر اسئلة محددة حول كيفية اداء المتعلمين على زوج معين من الفقرات، وتتطلب الاسئلة المختلفة معاملات احصائية مختلفة، ومن غير المرغوب به تطبيق واحد أو اكثر من هذه المعاملات للازواج المحتملة من الفقرات جميعها في اختبار كبير وتؤدي الى معلومات مهمة. فلمجموعات الفقرات الصغيرة (مثل فقرات كتبت من الخصائص نفسها او المطابقة للهدف نفسه في دراسة صدق المحتوى). فقد تزودنا معلومات الاتفاق بأدلة تجريبية تدعم الحكم المنطقي لتشابه الفقرات وفي بعض الاختبارات المحكية تدون اداءات المفحوصين منفصلة لعدد من الاهداف المختلفة أو مجالات المهارات وبعدد قليل من الفقرات يقيس كل هدف، وفي مثل هذه الحالات يبدو ان التحقق من درجة الاتفاق في الاداء على الفقرات ضمن هدف يقيس مهارة مؤكدة يلزم لدعم تفسير الدرجات.

ملاحظة اخيرة، وتطبق على طرائق تحليل الفقرات المحكية جميعها، لاحظ كل من هاريس (Harris , 1974) وهاريس وبيرلمان وويلكوكس (harris , Pearlman & Wilcox , 1977) اخذوا موقف في حالة اختيار الفقرات عشوائياً من ملف اختباري لتمثيل نطاق محتوى اختير بعناية، فلا يجب حذف اي فقرة بالاعتماد على تحليل الفقرات. فقد اكدوا على ان المراجعة بعناية للفقرات اثناء مرحلة بناء الاختبار والمناداة بخطة تطوير الاختبار تكون كافية للحصول على درجات اختبار صادقة. ويكون دور بيانات تحليل الفقرات التزويد بمعلومات حول ما إذا كان النظام التدريسي أو خطة تطوير الاختبار فاشلة، وتتضمن وجهة النظر هذه ان على كتبة الفقرات لاختبار ما تجنب انتاج فقرات فيها قصور أو استخلاص استدلالات عن المفحوصين من الفقرات السيئة . وانتقد هذا الموقف من قبل ميسك (Messick, 1975) وأدى الى طرح سؤال من قبل بيترسون (Petersen ,1979) حول منطقية الاعتمادعلى عينة من الفقرات في بناء الاختبار عندما يكون الهدف الاستدلال عن الاداء على نطاق فقرات اكبر، لذا يجب ان يكون مستخدم الاختبار مدركا بان الطرائق التي تستخدم فيها بيانات تحليل الفقرات يعتمد على وجهات نظر فلسفية مختلفة حول بناء الاختبار وتفسير درجاته.

الخلاصة:

تحليل الفقرات هوحساب وفحص أية خاصية احصائية لتوزيع الاستجابات على الفقرة وللفقرات ثنائية التصحيح فإن الوصف الاكثر شيوعاً هو صعوبة الفقرات (P) والذي يعبر عن نسبة المفحوصين الذين اجابوا على الفقرة إجابة صحيحة. وقيم P للفقرة ليست بالضرورة دليل أو معامل لنوعية الفقرة. ومع ذلك فللاختبارات معيارية المرجع فإن قيم P القريبة من (0.5) تؤدي الى اقصى تباين درجة كلي، وبالتالي أقصى ثبات. وعند استخدام فقرات الاختيار من متعدد يكون هنالك فرصة للتخمين، لذا فإن قيم P الاكبر من (0.5) تعد نموذجية لتعظيم ثبات الدرجة الكلية.

وتستخدم مؤشرات تمييز الفقرة لتقييم مدى ارتباط الاداء على الفقرة بمحك آخر وقد يكون هذا المحك الدرجة الكلية أو متغير آخر، وعندما تستخدم الدرجة الكلية فإن اختيار الفقرات التي تميز أكثر تؤدي الى اتساق داخلي أكبر

لدرجات الاختبار. وللفقرات غير ثنائية التصحيح يكون معامل ارتباط بونيت بايسيريال مفيداًفي الربط بين درجات الفقرة ودرجات المحك وعندما تكون الفقرات ثنائية التصحيح فقد يكون دليل التمييز(D) وارتباط بونيت بايسيريال أو ارتباط بونيت بايسيريال من المؤشرات المهمة، وذلك عندما تكون درجات المحك متصلة. وعندما تكون درجات المحك ثنائية التصحيح فإن معامل فاي او تتراشورك يكون مناسباً، والمقارنات بين هذه المعاملات الخمس تشير الى ان ارتباطها عالٍ على الاغلب، مع اكبر فروق تظهر للفقرات متطرفة الصعوبة. لذا فإن الاختيار عبر هذه الطرائق يجب ان يكون بناءاً على طبيعة المتغيرات ونوع المعلومات المرغوب فيها والاعتبارات الحسابية. كذلك تم عرض صيغ حساب الاخطاء المعيارية لهذه الاحصائيات ومؤشر ثبات الفقرة عبارة عن حاصل ضرب الانحراف المعياري للدرجة وارتباط الفقرة مع درجة الاختبار اكلي، ومعامل صدق الفقرة عبارة عن حاصل ضرب الانحراف المعياري للدرجات وارتباط درجات الفقرة مع المحك الخارجي. ويمكن حساب تباين درجة الاختبار الكلي ومعامل ارتباط الفا من معاملات الثبات ويمكن حساب معامل صدق الدرجة الكلية من معاملات ثبات الفقرة وصدقها.

وعلى الاغلب، يقرر مطوّر الاختبار في دراسة تحليل الفقرة اي خصائص درجات الاختبار الكلية هي الاكثر اهمية، وتحديد معاملات الفقرة التي لها اكبر الاثر على هذه الخصائص. وتطبق الفقرات على عينة مفحوصين (عادة 100 أو اكثر) وتحليل استجابات الفقرات، وتؤسس خطة الاختيار الفقرات التي لها المساهمة الأكبر في خصائص الاختبار (وكذلك لتحديد الفقرات التي تتضمن قصور، وبعد اختيار مجموعة الفقرات الفرعية تطبق على عينة أخرى مستقلة عن الأولى لدراسة الصدق التقاطعي (وفي العديد من الحالات ولتسهيل الدراسة فإن نصف العينة الاستطلاعية تستخدم في دراسة الصدق التقاطعي).

وللاختبارات محكية المرجع، قد لا تكون معاملات تحليل الفقرة التقليدي مناسبة لأن معظم احصائيات تمييز الفقرة صممت لاختيار الفقرات التي تعظم التباين عبر المفحوصين، وهذا هدف أكثر ملائمة بدرجة أكبر للاختبار معياري المرجع،

ومع ذلك فإن مطوروا الاختبارات محكية المرجع قد يهتموا بصعوبات الفقرة وحساسية الفقرة للتعليم ودرجة الاتفاق بين ازواج معينة من الفقرات، واقترحت طرائق احصائية معنية في ادبيات القياس لفحص هذه الأنواع من الخصائص التي عرضت هنا. وأخيراً من المهم ملاحظة أن اختيار الفقرات على أساس بيانات تحليل الفقرة تكون أقل ملائمة للاختبارات المحكية عنه للاختبارات المعيارية، إذ أن تمثيل النطاق قد ينخفض لمثل هذا الاختيار بناءاً على المحك الاحصائي، وحتى للاختبارات المعيارية فإن طبيعة النباء المراد قياسه قد تتبدل فيما لو تم اختيار الفقرات على أساس محك إحصائي صرف دون ردّها لمواصفات الاختبار الأولى.

التمارين:

1/ استخدم بيانات الفقرتين الاتيتين لعشرة مفحوصين في الإجابة على الاسئلة التي تلي:

المفحوص	الفقرة 1	الفقرة 2	الدرجة الكلية
1	0	1	12
2	0	1	15
3	1	1	16
4	0	0	10
5	1	1	7
6	1	0	5
7	1	0	6
8	0	1	10
9	1	1	15
10	1	0	13

أ- ما معاملات التمييز للفقرتين اعتماداً على تجزئة 50/50 ؟ وما قيمة المعاملات لتجزئة 27% للفئتين العليا والدنيا؟

ب- ما قيم معاملات بوينت بايسيريال $\hat{\rho}$ لهاتين الفقرتين؟

ج- ما قيم معاملات بوينت بايسيريال النقطي $\hat{\rho}$ لهاتين الفقرتين؟

د- كيف تختلف الافتراضات المتضمنة في معاملات بايسيريال و بوينت بايسيريال؟

2/ تفحص معلومات تحليل الفقرة الآتية وأجب عن الأسئلة التي تلي:

رقم الفقرة	استجابات الفقرة % 1	2	3	4	اخرى	صعوبة الفقرة	تمييز الفقرة	ارتباط بوينت بايسيريال
21	11	5	149+	11	0	0.84	0.15	0.23
22	34	11	127+	4	0	0.72	0.30	0.27
23	11	4	32+	128	1	0.18	0.13-	0.13-
24	30	34	1	111+	0	0.63	0.64	0.52
25	6	0	20	150+	0	0.85	0.34	0.34

أ- أي الفقرات يبدو أنها الأكثر حاجة للمراجعة؟

ب- للفقرتين 21 و 25 معاملات صعوبة متساوية تقريباً، ومع ذلك فإنها تختلف في تمييزها، فكيف يمكن تفسير هذا؟

ج- إذا أردت تحسين الفقرة 21، فكيف تحاول ذلك؟

د- اقترح على الأقل فرضيتين معقولتين لتفسر التمييز السلبي للفقرة 23؟

هـ- احسب قيم معامل ارتباط بايسيريال بين الفقرة والدرجة الكلية للفقرتين 24، 25؟

و- ما أقل قيمة مقبولة لارتباط بوينت بايسيريال إذا أراد مطور الاختبار أن يكون متأكداً بنسبة 95% بأن هذا الاحصائي أكبر من صفر وبدلالة؟

ز- افترض نموذج تخمين عشوائي، ما مستوى الصعوبة المثالي لهذه الفقرات إذا رغب مطور الاختبار تعظيم ثبات درجات الاختبار؟

3. افترض أن الفقرتين (21 و 22) في المثال السابق أعدت على وفق مواصفات الفقرات نفسها وكان توزيع الاستجابات الثنائي لهما على النحو الآتي:

	الفقرة 21	
الفقرة 22	+	−
+	120	7
−	29	20

أ– ما الطريقة الاحصائية التي يمكن استخدامها لاختبار فرضية أن الفقرتين (21 و 22) تقيسان المهارة أو المعرفة نفسها ؟

ب– ما الطريقة الاحصائية التي يمكن استخدامها لاختبار فرضية أن الفقرة (22) أكثر صعوبة من الفقرة (21)؟ أجري الحسابات ووضح استنتاجك؟

ج– فسر كيف استخدم الاحصائي في تمرين (3–أ) ليعُزى إلى معامل فاي، وهل أن هذين المعاملين يزودان بالمعلومات نفسها؟

4/ بين أي من احصائيات تحليل الفقرة أكثر ملائمة لكل موقف، استخدم الدليل الآتي لبيين استجابتك:

1. معامل فاي بين درجات الفقرة.

2. معامل ارتباط بوينت بايسيريال بين الفقرة والدرجة الكلية.

3. ارتباط بوينت بايسيريال بين الفقرة ودرجة محك خارجي.

4. ارتباط بيرسون بين الفقرة والدرجة الكلية.

5. ارتباط بيرسون بين الفقرة والدرجة على محك خارجي.

أ– مرشد في مركز ارشادي بكلية يرغب في اختيار الفقرات الصادقة في مقياس اتجاهات لاستخدامها في التنبؤ بسلوكات الطلبة الذكور في الكلية إذ صححت كل فقرة على تدريج تراوح بين (1 و 7)، وكان المحك الارتباط مع الاخاء الاجتماعي (وصنف الطلبة على أنهم أعضاء أو غير اعضاء).

ب– رغب مشرف على اختبارات تطوير ثبات اختبار إعادة لقائمة سلوكات حيث صححت الفقرات بتدريج ثنائي (1) للإجابة نعم و (0) للإجابة لا.

ج– يرغب استاذ جامعي زيادة الاتساق الداخلي لاختبار وحدة مؤلف من (50) فقرة من

نوع الاختيار من متعدد وبأربعة بدائل لكل فقرة / وأعطيت نقطة واحدة لكل فقرة اجابتها صحيحة.

د- يرغب عالم نفس تدريج فقرات على مقياس ليكرت التي تتأثر استجابات الأفراد على الفقرات باعمارهم.

هـ- طوّر عالم نفس قائمة لتقييم الاحكام العرقية وصححت كل فقرة على تدريج مؤلف من (5) نقاط، وتم اعدادها من خلال مساهمات خاصة على اعضاء مقطع عرضي واسع للمجموعة الثانوية ولم يستطع عالم النفس تحديد سلوك واحد (سهل الملاحظة ضمن حدود الوقت) ويمكن استخدامه كمحك.

5/ أ- اشتق صيغة لمعامل ارتباط بوينت بايسيريال ابتداءاً من صيغة معامل ارتباط بيرسون:

$$P_{xy} = \frac{\sum (x - \mu_x)(y - \mu_y)}{N\sigma_x \sigma_y}$$

حيث x تمثل متغير ثنائي يأخذ الدرجتين (1و 2).

ب - في ضوء هذا الاشتقاق، فسر لماذا تكون قيمة معامل بوينت بيسيريال = 1 نادرة الحدوث.

6/ أشار بعض المؤلفين أنه من أجل أن يكون للدرجات على أي مقياس مؤشر لصدق المحتوى أو البناء يجب أن تتوافر مؤشرات بأن كل فقرة ترتبط بالدرجة الكلية للاختبار، وحيث أن ارتباط درجة الفقرة بالدرجة الكلية ترتبط ارتباطاً مباشراً بالاتساق الداخلي، فما الأساس المنطقي لموقف المؤلفين هذا؟

7/ صمم طالب دراسات عليا اطروحته التي تتطلب قياس اتجاهات الامهات نحو المواليد الجدد الذين فيهم عيب ولادي معين، وسيحصل الطالب على العينة من مستشفيات المنطقة كلها، ومع ذلك كانت عينته صغيرة (تراوحت ما بين 25 إلى 30 طفل) خلال سنة.

وطرحت الكلية من خلال طاقم مشرفين أربعة مقترحات منفصلة، وعلى النحو الآتي:

1. استخدام أول (15) حالة في تحليل الفقرات ثم تنقية الأداة واستخدام الحالات الخمسة عشر الأخرى في الدراسة نفسها.

2- استخدم الأداة دون أي تحليل للفقرات، وافترض أن هنالك فرصة للحصول على معلومات جيدة.

3- طبق الفقرات جميعها على الأفراد جميعاً، وأجرى تحليل فقرات واحذف الفقرات السيئة ثم أعد تصحيح الأداة للأفراد أنفسهم باستخدام الفقرات الجيدة فقط.

4- لا تجري الدراسة.

ناقش مزايا ومساوىء كل مقترح وقرر أيها أفضل.

8/ افترض أن مطور الاختبار اعتبر أن إنتاج اختبار فرعي تألف من (15) فقرة التي عرضت بيانات تحليل فقراتها في جدول (14-4) . احسب التباين التقريبي وثبات الدرجات على هذا الاختبار الفرعي.

الفصل الخامس عشر

15

مدخل الى نظرية الاستجابة للفقرة

الفصل الخامس عشر

مدخل الى نظرية الاستجابة للفقرة

أشرنا في الفصل السابق إلى أن اختيار الفقرة يعتمد على معاملي الصعوبة والتمييز، وتعمل هذه بفاعلية في مواقف عدة في بناء الاختبارات، إلا أن تطوير الاختبار يتطلب معلومات إضافية عن الاستجابات للفقرة، افترض الفقرتين الاتيتين مع نماذج الاستجابة الخاصة بها وعلى النحو الآتي:

1. أجاب على الفقرة (أ) إجابة صحيحة كل من حصل على درجة 50 أو أكثر في الاختبار، وأجاب عليها بشكل خاطئ جميع المفحوصين الذين حصلوا على درجة أقل من 50.
2. أجاب على الفقرة (ب) إجابة صحيحة 20 % من الذين حصلوا على درجة 45، و 40% من الذين حصلوا على الدرجة 50، و60% من الذين حصلوا على الدرجة 55، و 80 % من الذين حصلوا على الدرجة 60، والمفحوصين جميعهم الذين حصلوا على الدرجة (65) فما فوق.

من الواضح وجود اختلافات في أنماط الاستجابة الملاحظة لكلا الفقرتين، ومن الضروري لمطوِّر الاختبار أن يعرف هذا، ولا يوفر التحليل الإحصائي التقليدي للفقرة معلومات عن كيفية الأداء على الفقرة من قبل مفحوصين بمستويات قدرة مختلفة للسمة، واحدى طرائق تطوير الاختبار التي تؤدي إلى صورة أكثر اكتمالاً لوظيفة الفقرة تُعرف باسم نظرية الاستجابة للفقرة أو نظرية السمة الكامنة.

وفي نظرية السمة الكامنة يفترض مطوِّر الاختبار أن الاستجابة على فقرات الاختبار تعزى إلى سمات كامنة تكون في عددها أقل من عدد فقرات الاختبار، وعلى العكس فأن معظم تطبيقات النظرية تفترض وجود سمة كامنة واحدة تفسر الاستجابة على فقرات الاختبار، وفي صميم النظرية نموذج رياضي يبين كيفية استجابة مفحوصين من مستويات مختلفة للقدرة على الفقرات، وهذه المعرفة تسمح بمقارنة أداء المفحوصين الذي يتقدمون لاختبارات مختلفة، كذلك تسمح بامكانية تطبيق نتائج تحليل الفقرة على مجموعات بمستويات قدرة تختلف عن المجموعة التي حللت الفقرات من خلال بياناتها، ومن الطبيعي أن مثل هذه الامكانات لها تطبيقات سيتم وصفها في هذا الفصل وفي الفصل العشرين، ومن المصادر المهمة للتطبيقات

العملية للنظرية:

(Hambelton, 1983) , (Hambelton & Swaminathan 1985) , (Traub & Wolfe, 1981) , Lord, (1980), (Drasgow & parsons, 1983).

وخصص هذا الفصل لعرض المفاهيم الأساسية وافتراضات النظرية، وعرض لأربع من نماذجها، وقد عرض لأحد هذه النماذج – النموذج الوغاريتمي الطبيعي – العلاقة بين معالم نظرية الاستجابة للفقرة ومعالم النظرية التقليدية، كذلك تم عرض العديد من تطبيقات النظرية وقبل البدء بنماذج السمة الكامنة وتطبيقاتها من الضروري عرض بعض المفاهيم الأساسية وتحديداً يجب على القارئ استيعاب المقصود بكل من:

1. ماذا نعني بالسمة الكامنة.
2. مم يتألف المنحنى المميز للفقرة.
3. ما هو فرض الاستقلال المحلي وكيفية ارتباطه بابعاد الاختبار.

وهذه المفاهيم الثلاثة أساسية في نظرية الاستجابة للفقرة (السمة الكامنة)، وسيتم وصفها في الأجزاء التالية.

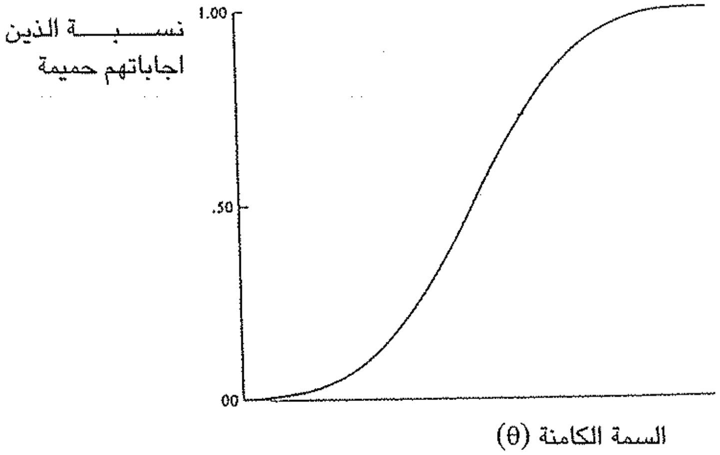
المفاهيم الأساسية في نظرية الاستجابة للفقرة:

السمات الكامنة والمنحنيات المميزة للفقرة

أحد المفاهيم المركزية لنظرية الاستجابة للفقرة هو المنحنى المميز للفقرة (Item characteristic Curve ICC)، ويمثل احتمالية الاستجابة الصحيحة على الفقرة كدالة للسمة الكامنة (ويرمز لها بالرمز θ) المقاسة في فقرات الاختبار (وفي الجزء التالي سنبين بدقة أكثر ماذا نقصد عندما نقول السمة الكامنة المتضمنة في الأداء على فقرات الاختبار)، ففي معظم تطبيقات نظرية الاستجابة للفقرة نفترض أن المنحنى المميز للفقرة له الشكل S كما هو مبين في الشكل (15–1) ويبين المنحنى أنه بزيادة قيمة السمة الكامنة تزداد احتمالية الاستجابة الصحيحة على الفقرة، وتكمن أهمية المنحنى المميز للفقرة مقارنة باحصائيات الصعوبة والتمييز التي نوقشت في الفصل الرابع عشر أنها تسمح برؤية أن احتمالية الاستجابة الصحيحة على الفقرة تعتمد على درجة السمة درجة الكامنة.

ومن الأهمية بمكان التفسير الصحيح لاحتمالية الاستجابة الصحيحة على الفقرة، وقدم لورد (Lord, 1980) تفسيرين مقبولين لهذه الاحتمالية، فللتفسير الأول يجب تسمية مجموعة

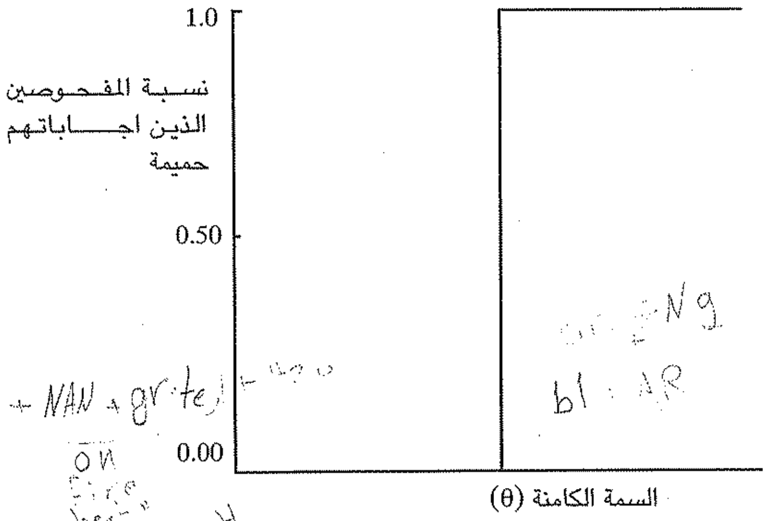
فرعية من المفحوصين عند كل نقطة على تدريج السمة الكامنة، والخصائص المحددة لكل مجموعة فرعية هي الأفراد الذين لديهم درجة السمة الكامنة نفسها (وفي الجزء المتبقي من هذا الفصل فإن أفراد المجموعة الفرعية سنطلق عليهم اسم المجموعة الفرعية المتجانسة)، وتفسر احتمالية الإجابة الصحيحة عندئذ على أنها احتمالية الإجابة الصحيحة لمفحوص اختير عشوائياً من المجموعة الفرعية المتجانسة، افترض أن المنحنى المميز للفقرة يشير إلى أنه عندما تكون θ =2 فإن احتمالية الإجابة الصحيحة على الفقرة تكون (0.87)، ويفسر هذا ليعني أن احتمالية الإجابة الصحيحة لمفحوص اختير عشوائياً من المجموعة الفرعية المتجانسة يساوي (0.87)، ويشير التفسير الثاني إلى المجموعة الفرعية من الفقرات التي لها المنحنى المميز للفقرة نفسه، وتفسر احتمالية الإجابة الصحيحة عندئذ على أنها احتمالية استجابة مفحوص معين على فقرة اختيرت عشوائياً من مجموعة الفقرات الفرعية، وأوصى لورد (1980) بالتحديد تجنب تفسير احتمالية الإجابة الصحيحة باحتمالية اجابة مفحوص معين على فقرة محددة إجابة صحيحة، وسنستخدم في الكتاب التفسير الأول لتجنب التداخل عبر التفسيرات الثلاث، وسنبحث المنحنى المميز للفقرة على أنه تمثيل بياني لنسبة المفحوصين الذين أجابتهم صحيحة على الفقرة وعلى أنها دالة لدرجة السمة الكامنة.



شكل (15-1) العلاقة بين القدرة والاستجابة على الفقرة

ومع ذلك فإن الشكل S للمنحنى المميز للفقرة يستخدم بشكل واسع الانتشار في بناء الاختبارات، وهو ليس الشكل الوحيد المحتمل للمنحنى المميز للفقرة، وسنبحث في المناقشة الحالية منحنيات مثل تلك التي تظهر في الشكل (15-2)، ويطلق على مثل هذه المنحنيات اسم

دالة الخطوة (Step function) ، ويتضمن هذا المنحنى وجود درجة سمة كامنة دنيا يرمز لها بالرمز (θ^2) وعند درجة أقل منها لا يستطيع المفحوص الإجابة بشكل صحيح على الفقرة، وتفيد دوال الخطوة هذه في إنتاج مفاهيم مهمة عديدة لنظرية الاستجابة للفقرة، ومع ذلك تعد دوال الخطوة للمنحنى المميز للفقرة أقل شيوعاً في بناء الاختبارات من منحنيات الشكل S، وذلك لأن بيانات الاختبار الحقيقية تكون أكثر اتساقاً في حالة المنحنى ذو الشكل S.



شكل (15-2) : دالة الخطوة للمنحني المميز للفقرة

أحادية البعد والاستقلال المحلي:

في نظرية القياس تدرس العلاقة بين متغيرين من خلال الارتباط بينهما، ومن المألوف في نظرية الاستجابة للفقرة استخدام مفاهيم أكثر عمومية للاستقلال الاحصائي أو عدمه عند الحديث عن العلاقة بين المتغيرات، ويمكن تحديد المفاهيم للفقرات ثنائية التدرج على النحو الآتي:

لتكن P_i (+) ترمز إلى احتمالية الإجابة الصحيحة على الفقرة (i)، و P_i(−)، ترمز إلى احتمالية الإجابة الخاطئة على الفقرة (i)، وكذلك P_j (+) و P_j (−) ترمز إلى الاحتمالات المناظرة للفقرة (j)، ولترمز (P+,+) و P(−,+) و P(− , −) إلى احتمالات نماذج الإجابة المبينة في الأقواس، فعلى سبيل المثال ترمز P (+,+) إلى احتمالية الإجابة الصحيحة على الفقرتين i , j على التوالي، وباستخدام هذه الرموز فإن الفقرتين تكونان مستقلتين احصائياً إذا كان:

$$P_i(+) \cdot P_j(+) \quad = P(+,+) \qquad (15-1-أ)$$

و

$$P_i(+) \cdot P_j(-) \quad = P(+,-) \qquad (15-1-ب)$$

و

$$P_i(-) \cdot P_j(+) \quad = P(-,+) \qquad (15-1-جـ)$$

و

$$P_i(-) \cdot P_j(-) \quad = P(-,-) \qquad (15-1-د)$$

وتكون الفقرات غير مستقلة احصائياً في حالة عدم تحقق احدى معادلات (15-1)، فعلى سبيل المثال إذا كانت $P_i(+)=0.8$ و $P_i(-) = 0.2$ و $P_j(+) = 0.6$ و $P_i(-)= 0.4$ ، فإن استقلال الفقرتين يتحقق إذا كان:

$0.48 = P(+,+)$

$0.33 = P(+,-)$

$0.12 = P(-,+)$

و $0.8 = P(-,-)$

بمعنى أخر، تبين الشروط في المعادلة (15–1) أن الفقرات تكون مستقلة عندما يمكننا حساب احتمالية أي نمط استجابة لكلا الفقرتين من معرفة احتمالية الاستجابة الصحيحة والخاطئة فقط ولكل فقرة.

وتتحدد احادية البعد من خلال الاعتماد الاحصائي عبر الفقرات، وتحديداً فإن متطلبات احادية البعد للاختبار تعني أن الاعتماد الاحصائي عبر الفقرات يعزى إلى سمة واحدة فقط، وهذا يعني أن الاختبار احادي البعد إذا كانت الاختبارات احصائياً في المجتمع الكلي تقيس سمة كامنة واحدة فقط، ومثل هذه الفقرات تكون مستقلة احصائياً لكل مجموعة فرعية من المفحوصين المتجانسين بالنسبة للسمة الكامنة، وتتحدد هذه الاستقلالية بوساطة مجتمع فرعي من المفحوصين يقعون على نقطة واحدة على متصل السمة الكامنة، وهذا ما يطلق عليه اسم الاستقلال المحلي.

وهنا نشير إلى نقطتين مهمتين: الأولى هي أن الاستقلال المحلي واحادية البعد ليسا مفهومين متكافئين، فقد يكون الاختبار ثنائي البعد إذا وجد أن سمتين متكافئتين لمجموعة

مفحوصين متجانسين على كلا السمتين وفقراتهما مستقلة محلياً، وبشكل عام فإن ابعاد الاختبار يساوي عدد السمات الكامنة اللازمة لتحقيق الاستقلال المحلي.

والثانية: أنه لا يمكننا القول بالتأكيد وجود سمة واحدة فقط (أو أي عدد اخر منها) بحيث تكون الفقرات مستقلة محلياً، لذا فإن الاستقلال المحلي وعدد السمات الكامنة افتراضات اصطلاحية، ومع ذلك فإنه يمكن التحقق من صحة هذه الافتراضات، وسنقدم المراجع المناسبة لهذا الموضوع والمذكورة في أدبيات القياس.

وفي البدء، اقترحنا معنى دقيق لقولنا بأن سمة كامنة تفسر الأداء على فقرات الاختبار، وببساطة فإن هذا يعني أن السمة الكامنة تفسر الاعتماد الاحصائي عبر الفقرات، كذلك فليس من الضروري أن تكون السمة الكامنة قياس احصائي صادق للبناء الذي تهدف فقرات الاختبار قياسه، وهذا يعني أن الفقرات قياسها ضعيف للبناء، وعلى هذا فإنه يجب أن لا نساوي بين فكرة السمة الكامنة بالفكرة الأكثر عمومية للبناء.

ولتوضيح مفاهيم أحادية البعد والاستقلال المحلي سنعرض مثال افتراضي لعينة مفحوصين عددهم (200) تقدموا لاختبار مؤلف من أربعة فقرات، ومع ان هذا لا يشبه موقف الاختبار الحقيقي إلا أنه يمكن افتراض نماذج الاستجابات الخمسة الملاحظة كما هي مبنية في الجدول (15-1)، وأن استجابات (40) مفحوص تطابق أحد النماذج الخمسة، ويقال لنماذج الاستجابات الخمس هذه أنها تشكل تدريج جوتمان التام (والتداريج من هذا النمط موضحة باختصار في الفصل الثالث).

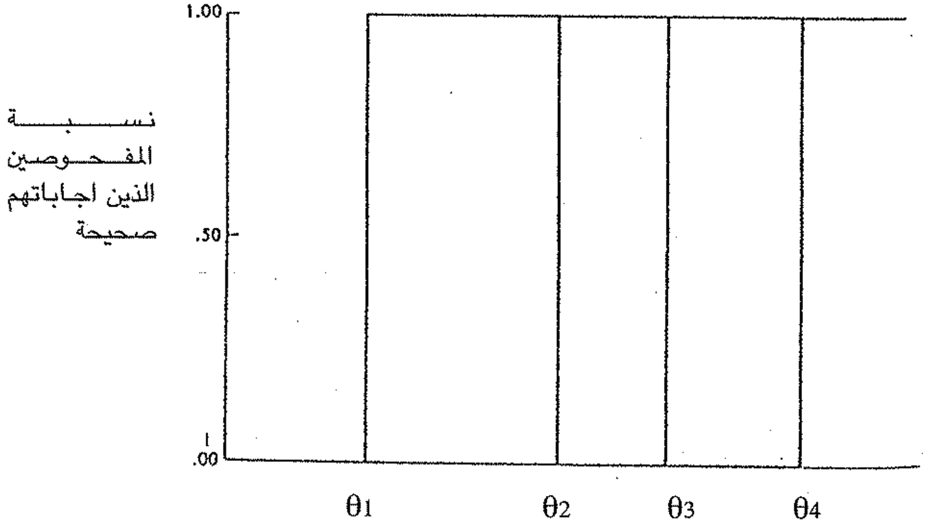
جدول (15-1) أنماط الاستجابة لاختبار افتراضي مؤلف من أربع فقرات

نمط الاستجابة	الفقـرة			
	1	2	3	4
أ	-	-	-	-
ب	+	-	-	-
ج	+	+	-	-
د	+	+	+	-
هـ	+	+	+	+

تؤشر-1 إلى استجابة خاطئة و (+) إلى استجابة صحيحة

وتتسق أنماط الاستجابة في جدول (15-1) مع فكرة أن استجابات الفقرات جميعها تعتمد على سمة كامنة واحدة، وأن المنحنى المميز لكل فقرة عبارة عن دالة خطوة، ويبين الشكل (15-3) دالة الخطوة للمنحنى المميز للفقرات الأربعة، وتمثل الرموز (θ_1، θ_2، θ_3، θ_4) درجات السمة الكامنة مرتبة تصاعدياً، فالمفحوص الذي قدرته الكامنة أقل مرتبة تصاعدياً، فالمفحوص الذي قدرته الكامنة أقل من θ_1 تكون إجاباته خاطئة على الفقرات الأربعة جميعها، وبالتالي فإن نموذج استجابته هو (– ، – ، – ، –).

والمفحوص الذي قدرته الكامنة تقع بين θ_1 و θ_2 يجيب إجابة صحيحة على الفقرة الأولى فقط، وبالتالي فإن نموذج استجابته يكون (+ ، – ، – ، –)، ويبين المحور الأفقي للشكل (15-3) نماذج الاستجابة للنقاط الأخرى على تدريج السمة الكامنة.



شكل (15-3): دالة الخطوة للمنحنيات المميزة للاختبار المؤلف من 4 فقرات

جدول (15-2): التوزيع المشترك لاستجابات المفحوصين جميعهم

الفقرة الرابعة

الفقرة الثالثة		+	–	
	+	0.20	0.20	0.4
	–	0.00	0.60	0.6

لنركز الآن على الفقرتين الثالثة والرابعة، ويبين جدول (15–2) احتمالية أنماط الاستجابة الممكنة على هذه الفقرات، وهذه الاحتمالات للمفحوصين جميعهم وعددهم (200) مفحوص، وبالتعريف فإن الاستجابات على هذه الفقرات مستقلة احصائياً إذا وفقط إذا تحققت معادلات (15–1) الأربعة جميعها، وإذا استخدمنا الاحتمالات المبينة في جدول (15–2) فإنه من الواضح عدم تحقق هذا لهاتين الفقرتين فمثلاً (+ و +) P= 0.20 ، بينما قيمة P3(+) . P4(+) =0.2 x 0.4= 0.08، لذا فإن (+ و +) P ≠ P3 P4(+)، واحصائياً فإن الفقرتين غير مستقلتين احصائياً (أي محلياً) عن المجتمع الكلي للمفحوصين.

ويمكننا القول نتيجة لذلك أن هاتين الفقرتين تشتركان في مصدر تباين، وبإجراء العملية نفسها لكل زوج من الفقرات يمكن أن نبين أن استجابات الفقرة تعتمد عى الفقرات الأخرى في الاختبار.

إن متطلب احادية البعد يعني وجود سمة كامنة واحدة لكي يكون زوج الفقرات مستقل محلياً، وهذا يعني الاستقلال الاحصائي لفقرات أي مجتمع فرعي متجانس على السمة الكامنة، وسنبدأ باختبار ما إذا كانت الفقرتين الثالثة والرابعة مستقلتين محلياً لفئة المفحوصين للقدرة الكامنة (θ_3) (وهؤلاء عبارة عن المفحوصين ذوي نمط الاستجابة (د) في جدول (15–1))، ويبين جدول (15–3) احتمالية كل نمط استجابة للفقرتين الثالثة والرابعة لمجتمع المفحوصين الفرعي ذوو القدرة الكامنة (θ_3)، ولكي تكون استجاباتهم مستقلة احصائياً يجب أن تتحقق العلاقات:

$$P_i(+/\theta_3) \,.\, P_j(+/\theta_3) = P(+, +/\theta_3) \qquad (15\text{-}2\text{-أ})$$

$$P_i(+/\theta_3) \,.\, P_j(-/\theta_3) = P(+, -/\theta_3) \qquad (15\text{-}2\text{-ب})$$

$$P_i(-/\theta_3) \,.\, P_j(+/\theta_3) = P(-, +/\theta_3) \qquad (15\text{-}2\text{-جـ})$$

$$P_i(-/\theta_3) \,.\, P_j(-/\theta_3) = P(-, -/\theta_3) \qquad (15\text{-}2\text{-د})$$

جدول (15–3): التوزيع المشترك لاستجابات المجموعة الفرعية عند $\theta_3=\theta$

الفقرة 4

	–	+	الفقرة 3
1.00	1.00	0.00	+
.00	.00	0.00	–
	1.00	0.00	

وتشير الرموز إلى الاحتمالات الممكنة جميعها عند θ_3 على متصل السمة الكامنة، ومن الاحتمالات المبنية في جدول (15-3) يتبين أن الاحتمالات الأربعة الممكنة تتحقق، ولهذا فإن الفقرتين الثالثة والرابعة مستقلتين احصائياً للمفحوصين الذين سمتهم الكامنة $\theta = \theta_3$، وثانية، بإجراء هذه العملية لكل زوج من الفقرات وعند كل نقطة على متصل السمة الكامنة يمكننا تبيين أن الاستجابات على ازواج الفقرات جميعها مستقل احصائيات عندما تبقى درجة السمة الكامنة ثابتة، وعلى هذا فإن الفقرات مستقلة احصائياً.

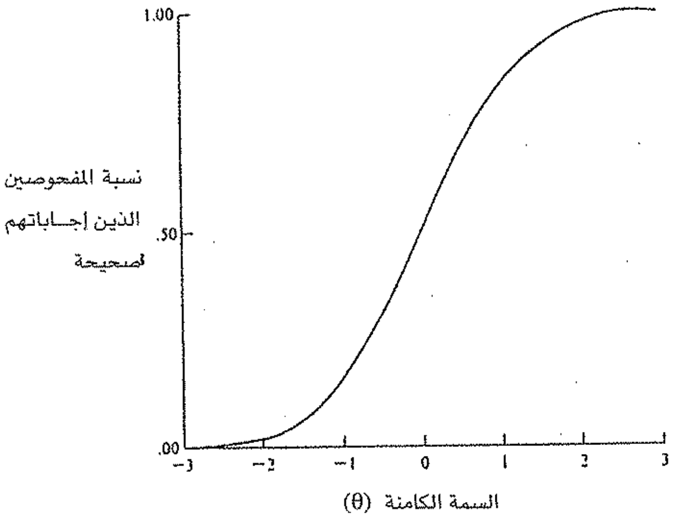
وكما سنبين فيما بعد فإن الاستجابات على الفقرة غير مستقلة للمجتمع الكلي، وبالتالي فإنها تشترك في واحد أو أكثر من مصادر التباين المشترك، كما سنبين أيضاً وجود السمة الكامنة بحيث أن المفحوصين المتجانسين بالنسبة للسمة الكامنة تكون مستقلة محلياً، وفي هذه الحالة يفسر التباين المشترك بين الفقرات ككل بالسمة الكامنة الوحيدة، وكنتيجة فإن الاختبار يحقق احادية البعد.

القياس المتحرر من عينة التطبيق:

النقطة المهمة في نماذج السمة الكامنة أنها تسمح بمقارنة المفحوصين حتى عندما يتقدمون لاختبارات مختلفة، ويشار إلى هذا أحياناً بأن القياس متحرر من عينة التطبيق. ولتوضيح هذه الخاصية افترض أن الفاحص يعرف بأن الفقرات الأربعة لها دالة الخطوة ICC الموضحة في شكل (15-3)، وهنا يختبر أحد المفحوصين بالفقرتين الأسهل، والمفحوص الأخر يختبر بالفقرتين الأصعب، وأجاب المفحوص الأول على الفقرة الأولى إجابة صحيحة وعلى الفقرة الثانية إجابة خاطئة، لذا فإن قدرة هذا المفحوص تقع في الفترة $\theta_1 \leq \theta \leq \theta_2$ ، بينما أجاب المفحوص الأخر على الفقرة الثالثة اجابة صحيحة وعلى الفقرة الرابعة اجابة خاطئة، استنتج الباحث أن القدرة الكامنة للمفحوص تقع في الفترة $\theta_3 \leq \theta \leq \theta_4$ ، و يمكن أن يستنتج الفاحص أن قدرة المفحوص الأول أقل من قدرة المفحوص الثاني، ويوضح هذا المثال أنه وباستخدام نظرية السمة الكامنة يمكن وضع المفحوصين على التدريج نفسه حتى عندما يتقدمون لفقرات اختبارية مختلفة، وهذا يشير إلى أن الفقرات جميعها تقيس السمة الكامنة نفسها، والمقارنة الدقيقة بين هذين المفحوصين في هذا المثال غير ممكنة بسبب قصر الاختبار، وبوضوح فإنه إذا تم بناء اختبار بعدد اكبر من الفقرات التي لها دالة منحنى مميز يمكننا اجراء مقارنة أكثر دقة، وعلاوة على ذلك فإن إحلال المفحوصين الذين تقدموا لاختبارات مختلفة على تدريج مشترك يكون لإختبارات مؤلفة من فقرات تتميز بمنحنيات مميزة للفقرة ذات صيغ أخرى، وقد أجريت دراسات معادلة تستخدم نظرية السمة الكامنة وتعتمد على الاختبار كقياس متحرر من عينة التطبيق من قبل **(Rentz&Rashaw, 1977)**و **(Marco,1977)**.

المنحنى الطبيعي:

إن معظم تطبيقات نظرية السمة الكامنة تفترض أن المنحنى المميز للفقرة له الشكل S، وأحدها هو المنحنى الطبيعي والذي طبق بكثرة في الأعمال المبكرة لنظرية السمة الكامنة (انظر (Lawley, 1943)، و (Lord, 1952b, 1953) ويبين الشكل (15-4) خصائص مهمة عديدة للمنحنى المميز للفقرة للمنحنى الطبيعي:



شكل (15-4): المنحنى الطبيعي المميز للفقرة، متوسط القدرة = صفر والانحراف المعياري = 1.

1- يرتفع المنحنى باضطراد بالاتجاه من اليسار إلى اليمين، ومثل هذا المنحنى يقال أنه يزداد باضطراد.

2- يقترب الخط التقاربي الأدنى من الصفر ولكنه لا يصل إلى الصفر أبداً، ويصل الخط التقاربي الأعلى إلى 1.

3- يرتبط المنحنى الطبيعي مباشرة بالتوزيع الاعتدالي المطروح في الفصل الثاني، وتعلمنا أنه عندما نعبر عن الدرجات بالدرجات الزائية فإنه يمكننا استخدام جدول التوزيع الاعتدالي المعياري للحصول على المساحة تحت المنحى الطبيعي على يسار الدرجة الزائية، وهذه المساحة أقل من 1، لذا فإنه يمكن تفسيرها على أنها نسبة، وهذه النسب للمنحنى الطبيعي الاعتدالي عبارة عن دالة للدرجات الزائية.

وعند استخدام المنحنى الطبيعي على أنه منحنى مميز اللفقرة فإن القيم على المحور الأفقي

تمثل القيم الممكنة للسمة الكامنة θ، وارتفاع المنحنى عند أي من قيم θ تمثل نسبة المفحوصين عند مستوى القدرة هذا الذي يمكنهم الإجابة على الفقرة إجابة صحيحة.

ويمكن كتابة معادلة المنحنى المميز للفقرة للمنحنى الطبيعي على النحو الآتي:

$$\int_{-\omega}^{w} f(z)\,dz = P_g(\theta) \qquad (3\text{-}15)$$

حيث ترمز $Pg(\theta)$ إلى نسبة المفحوصين الذين قدرتهم الكامنة θ، وهؤلاء هم الذين اجابوا على الفقرة g إجابة صحيحة، والتعبير الذي يمثل التكامل في المعادلة يعني المساحة بين ω و α تحت المنحنى الطبيعي، و ω عدد حقيقي يتحدد بالمعادلة:

$$a(\theta - b) = w \qquad (4\text{-}15)$$

حيث ترمز a إلى معلم تمييز الفقرة، و b إلى معلم الصعوبة، (وهذه ليست معالم الصعوبة والتمييز في نظرية القياس التقليدية)، وتشبه ω أو القيمة المكافئة لها " $a(\theta\text{-}b)$ الدرجة الزائية " وذلك لأن $P(\theta)$ تمثل المساحة على يسار $a(\theta - b)$ لمنحنى التوزيع الاعتدالي المعياري، لذا فإن امتلك المفحوص درجة قدرة كامنة $\theta = 2$ ، وكانت قيم المعالم للمنحنى المميز للفقرة g هي: $a = 0.5$ و $b = 1.0$

فإن ω لهذا المفحوص تساوي:

$$a(\theta - b) = w$$

$$\cdot 5\,(2\text{-}1.0) = \cdot 5$$

وحيث أن المساحة على يسار 0.5 في المنحنى الاعتدالي = 0.69، فإن استخدام المعادلة (3–15) يشير إلى أن 0.69 من المفحوصين الذين سمتهم الكامنة $\theta = 2$ سيجيب اجابة صحيحة على الفقرة g ، لذا فإن توافر لدينا قيم كل من aو b للفقرة g ونريد معرفة نسبة المفحوصين عند قدرة كامنة معينة الذين يمكنهم الإجابة عن الفقرة اجابة صحيحة علينا:

1– حساب قيمة $a(\theta\text{-}b)$.

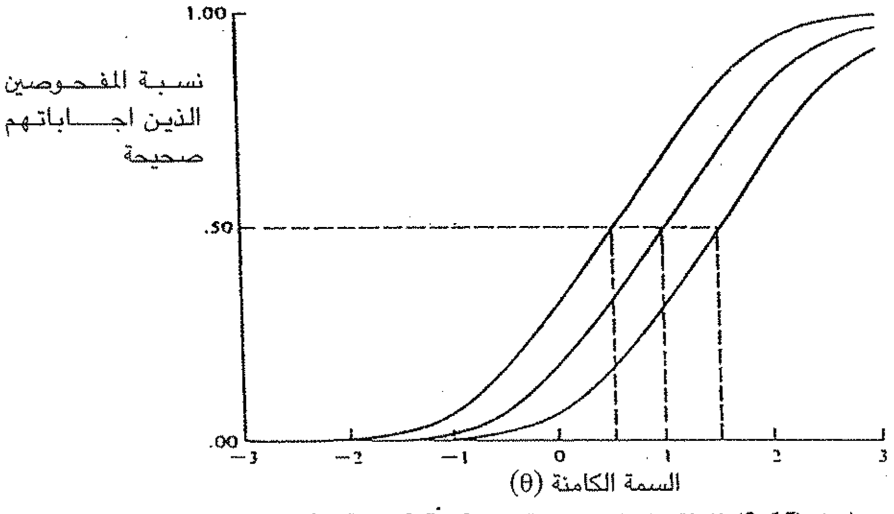
2– استخدام هذه القيمة لاستخراج $Pg(\theta)$ من جدول الدرجات الزائية الاعتدالي المعياري.

ومن المهم تذكر أن قيم المعالم b,a تتغير عبر فقرات الاختبار، لذلك فإن أدبيات القياس المتعلقة بالمنحنى المميز للفقرة ترمز للمعالم بالرموز bg , ag، حيث يدل الحرف g على الفقرة g في الاختبار، لهذا فإن معادلة المنحنى المميز للفقرة للمنحنى الطبيعي تكتب على النحو الآتي:

$$\int_{-\omega}^{a(\theta - bg)} f(z)\, dz = P_g(\theta) \qquad (5\text{-}15)\ldots\ldots\ldots$$

حيث أن ω= ag (θ-b_g)، ولنرى الآن كيف يعمل معلم الصعوبة b_g، دعنا نتعامل مع ثلاث فقرات لها قيم a_g نفسها وتختلف في قيم b_g، وهذه القيم هي: $0.5 = b_1$, $1.0=b_2$، $1.5=b_3$، ويبين الشكل (15–5) المنحنى المميز للفقرات الثلاث، وعلى كل منحنى خط عمودي نازل على متصل السمة الكامنة عند $P_g(\theta) = 0.5$، وبدءاً من اليسار فإن درجات السمة الكامنة تساوي 0.5، عند =b2 للفقرة الثانية وللفقرة الثالثة فأن (pg(θ=0 عند 1.0، 1.5 لاحظ أنه للفقرة الأولى أن $P_g(\theta)$= 0.5 عندما تكون $b_1 = \theta$، وبصورة مشابهة فإن $P_g(\theta) = 0.5$، عند $b_2=\theta$ للفقرة الثانية، وللفقرة الثالثة لذا فإن $P_g(\theta) = 0.5$، عند $b_3=\theta$ لذا فإن bg= درجة السمة الكامنة عندما يجيب 50 % من المفحوصين إجابة صحيحة على الفقرة g.

لنرى كيف يعمل معلم التمييز a_g، دعنا نتعامل مع ثلاث منحنيات مميزة للفقرات في الشكل 15–6، قيمة b لكل منها تساوي 1.5، ولكن قيم a هي: $0.1 = a_1$ ، $1 = a_2$ ، 100 = a_3 وعند النظر إلى الفقرة الأولى نرى أنها تميز بفعالية لأن نسبة المفحوصين الذين أجابوا على الفقرة اجابة صحيحة هو نفسه تقريباً عند درجات السمة الكامنة نفسها، كذلك فإن الفقرة 3 تميز بفعالية عالية بين المفحوصين الذين درجاتهم أقل من 1.45 وأعلى من 1.55، وفي الواقع فإن نسبة المفحوصين الذين أجابوا اجابة صحيحة أقل من 0.01 للذين سمتهم الكامنة θ أقل أو تساوي 1.45، في حين أنها أكبر من 0.99 للذين سمتهم الكامنة θ أكبر أو تساوي 1.55 .



شكل (15–2) ثلاثة منحنيات لها قيم $0.5 = b_1$ ، $1.0 = b_2$ ، $1.5 = b_3$

وأخيراً فإن الفقرة الثانية لها $a = 1$ تكون متوسطة من حيث تمييزها، لذا فإن a_g هو المعلم الذي يحدد ميل أو شدة انحدار المنحنى المميز للفقرة، فالفقرات غير الفعالة في تمييزها عند مستويات القدرة جميعها تكون ذات منحنى مميز مستو مع انخفاض قيم a_g، والفقرات التي تميز بشدة بين المفحوصين عند قدرة معينة (أعلى منها وأدنى منها) يكون المنحنى المميز لها شديد الانحدار مع ارتفاع قيم a_g، بينما الفقرات متوسطة التمييز عبر مدى واسع من القدرة يكون معدل انحدار المنحنى المميز لها معتدل مع اعتدال في قيم a_g أيضاً.

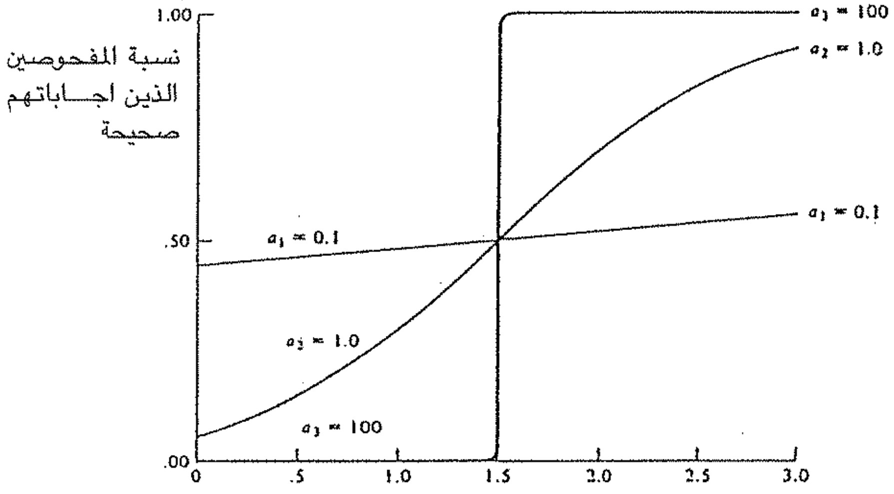
تدريج القياس:

إن تدريج متصل السمة الكامنة المبين في المعادلة (15-4) له نقطة أصل مناسبة، وكذلك الأمر بالنسبة لوحدة القياس، ويقصد بنقطة الأصل المناسبة أنه ولأي مجتمع فرعي متجانس يمكن أن يؤشر بالدرجة صفر لسمته الكامنة، إذ لا يمكن أن يتميز أي مجتمع فرعي بالغياب الكامل للسمة الكامنة، ويقصد بوحدة القياس المناسبة أنه بعد تعيين الدرجة الكامنة صفر لاحدى المجموعات فإنه يمكن لمجموعة أخرى قدرتها الكامنة أعلى من قدرة المجموعة الصفرية تعيين واحد على درجة السمة الكامنة لها، وفرق القدرة بين هاتين المجموعتين الفرعيتين هو وحدة القياس، ولأن كل من وحدة القياس ونقطة الأصل نقاط اعتبارية (اصطلاحية) فإنه يتم اختيار نقطة الأصل ووحدة القياس بحيث يكون متوسط السمة الكامنة صفر وانحرافها المعياري واحد لاحدى مجموعات القدرة أو السمة الكامنة، ونتيجة لذلك فمن الممكن ايجاد مخطط للمنحنى المميز للفقرة كما في الشكل (15-5)، وبتدريج يتضمن الأرقام الموجبة والسالبة على حد السواء.

وتعتمد قيم كل من ag و bgعلى الوحدة والتدريج المختارين لـ θ ، فإن تغيرت قيم الوحدة والتدريج لدرجات السمة الكامنة والتي يرمز لها بـ θ يكون مكافئ لتحويل θ وفق المعادلة:

$$K\theta + m = \theta^{\backslash} \qquad (6\text{-}15)$$

حيث أن كلاً من k , m اعداد حقيقية، و $\theta^{\backslash}$ درجة السمة الكامنة على التدريج الجديد، وعند تحويل θ إلى $\theta^{\backslash}$ فإن قيم كلاً من a_gو b_g تصبح على النحو الآتي:



شكل (15-6): ثلاثة منحنيات مميزة للفقرة لها قيم $a_1 = 0.1$ ، $a_2 = 0.1$ ، $a_3 = 100$

$$Kb_g + m = b_g^{\backslash} \quad \text{...........(15- 7 أ)}$$

$$K a_g + m = a_g^{\backslash} \quad \text{...........(15- 7ب)}$$

ويمكننا أن نبين بسهولة أن

$$a_g^{\backslash} (\theta^{\backslash} - b_g^{\backslash}) = a_g (\theta - b_g)$$

وبالتالي فإن $P_g(\theta)$ لا تتغير بتغير وحدة القياس وتدريجة.

ربط نظرية الاستجابة للفقرة بنظرية القياس التقليدية:

معالم المنحنى المميز للفقرة واحصائيات الفقرة التقليدية:

لتكن ρ_x تمثل معامل ارتباط بايسيريال بين درجات الفقرة g درجات السمة الكامنة، من الواضح أن هذا الارتباط مشابه لمعامل ارتباط بايسيريال بين الفقرة والدرجة الكلية والذي يستخدم ليمثل معامل التمييز عند تحليل الفقرات بنظرية القياس التقليدية، فإن كان توزيعθ توزيع اعتدالي متوسطة صفر وانحرافه المعياري واحد فإن:

$$a_g = \frac{\rho x}{\sqrt{1 - bg^2}}$$

جدول (15-4) المناظرة بين a_g و p_g

ag	Pg
0.10	0.1
0.20	0.2
0.31	0.3
0.44	0.4
0.58	0.5
0.75	0.6
0.98	0.7
1.33	0.8
2.06	0.9

لذا، فإن كان معامل ارتباط بايسيريال لفقرة ما مع السمة الكامنة $\rho_x = 0.30$ ، فإن قيمة معلم a للمنحنى المميز للفقرة يكون

$$a_g = \frac{0.30}{\sqrt{1 - (0.30)^2}} = 0.31$$

ويبين الجدول (15-4) قيم a_g المناظرة لعدة قيم b_g، ويتبين من الجدول أنه كلما اقتربت b_g من واحد فإن قيم a_g تزداد، لذا فمن البداهة أن مثل هذه العلاقة تدعم تفسير a_g على أنه معلم تمييز (معامل تمييز).

ويمكن ربط معلم الصعوبة للمنحنى المميز للفقرة b_g بمعاملات تحليل الفقرة التقليدي ايضاً، ويمكن ان نبين انه ووفق افتراضات التوزيع الاعتدالي أن:

$$bg = \frac{\phi^{-1}(- Pg)}{- Pg}$$

حيث ترمز p_g الى نسبة الاجابات الصحيحة لمعامل صعوبة الفقرة و $\phi^{-1}(P_g)$ الى الدرجة الزائية التي تقطع المساحة p_g الى يسار الدرجة الزائية في التوزيع الاعتدالي. ومن معرفتنا لقيم p_g يمكننا الحصول على قيمة $\phi(p_g)$ باستخدام جداول مثل الجدول المبين في الملحق أ . فعلى سبيل المثال اذا كانت $p_g = 0.84$ و $\phi^{-1}(P_g) = 1.00$، وكانت P_g للفقرة تساوي 0.30 فإنه يمكن حساب قيمة bg للفقرة على النحو:

$$b_g = \frac{-1.00}{0.30} = -3.33$$

المنحنيات المميزة للفقرة والثبات التقليدي:

يمكن التعبير عن الدرجة الحقيقية للاختبار وعلاقتها بـ θ بوساطة المنحنى المميز للفقرات، وكما يأتي:

$$T = \sum_g Pg(\theta) \qquad (10\text{-}15)...........$$

وهذه المعادلة صحيحة بغض النظر عن المنحنى المميز للفقرة ، وتبين ان العلاقة بين Tو θ ليست علاقة احصائية ، فالدرجة الحقيقية عبارة عن تحويل غير خطي للسمة الكامنة، ويبين جدول (15 –5) العلاقة بين T و θ لاختبار افتراضي مؤلف من ثلاث فقرات:

جدول (15–5) العلاقة بين السمة الكامنة والدرجات الحقيقية :

θ	$P1(\theta)$	$P2(\theta)$	$P_3(\theta)$	الدرجة الحقيقية
-2	0.022	0.105	0.382	0.51
-1	0.158	0.226	0.460	0.84
صفر	0.500	0.401	0.539	1.44
1	0.841	0.636	0.617	2.09
2	0.977	0.773	0.691	2.44

$1 = a_1 \quad 0 = b_1 \quad 0.5 = a_2 \quad 0.5 = b_2 \quad 0.2 = a_3 \quad -0.5 = b_2$

والعلاقة بين الدرجات الملاحظة ودرجات السمة الكامنة θ علاقة احصائية، ويعرف منحنى الانحدار للتنبؤ بـ x من درجات θ على انه المنحنى المميز للاختبار ومنحنى الانحدار هذا مشابه لمنحنى العلاقة بين T و θ لذا فإن معادلة X هي:

$$X = \sum_g P(\theta) + E \qquad (11\text{-}15)..............$$

و E هي خطأ القياس استنتاجاً من المعادلة (15-10)، ما مقدار القوة في العلاقة بين X و θ، وكيف يمكن تفسير قوة هذه العلاقة؟ في البداية، لان العلاقة بين X و θ غير خطية فان الانسب هنا استخدام معامل الارتباط غير الخطي:

$$\sqrt{1-\frac{\varepsilon\theta\,\sigma^2(x/\theta)}{\sigma_x^2}}=\eta_{x\,\theta} \qquad (12\text{-}15)................$$

وذلك لقياس قوة الارتباط بين X و θ ، اذ ترمز $\sigma^2(x/\theta)$ الى تباين X لمجتمع فرعي متجانس. ويمكن تبيان ان هذا التباين $\sigma^2(E/\theta)$ اي خطأ تباين المجتمع الفرعي مساوياً لتباين المجتمع الفرعي. كذلك يمكن تبيان ان $\varepsilon\sigma^2(x/\theta)$ مساويا لتباين الخطأ المحدد كلاسيكياً، ونتيجة ذلك يمكن اعادة كتابة المعادلة (15–12) على النحو الآتي:

$$\sqrt{1-\frac{\sigma^2 E}{\sigma_x^2}}=\eta_{x.\theta}$$

و يمكن رؤية ان هذا مساوياً لمعامل الثبات والمذكور في الفصل السادس. لذا فإن قوة العلاقة بين X و θ يساوي معامل ثبات الدرجات X.

النماذج اللوغاريتمية:

يعد المنحنى الطبيعي الصيغة السائدة للمنحنى المميز للفقرة في الابحاث الأولى لنظرية السمة الكامنة وقد حل اليوم مكانها نماذج لوغاريتمية ثلاثة تتطلب حسابات ابسط . وفي الاجزاء التالية سنقدم صيغة عامة لهذه النماذج، وبرنامج محوسب يستخدمان في حساب معالم هذه النماذج، وسنقدم مثال يبين المعلومات التي نحصل عليها من احد هذه البرامج.

تعد دالة المنحنى اللوغاريتمي التراكمي اساس المنحنى المميز للفقرة في هذه النماذج الثلاثة، وهذه الدالة لها الصيغة العامة الآتية :

$$\frac{e^{x}}{1+e^{x}}=P_g(\theta) \qquad (13\text{-}15)..............$$

حيث ترمز e الى الاساس للوغاريتم الطبيعي، و x هنا رمز اصطلاحي لا يقصد به الدرجة الملاحظة. ويعد المنحنى المميز للفقرة في النماذج الثلاثة تباين في الصيغة العامة للمعادلة (15 –13)، وتختلف هذه النماذج عن بعضها البعض في عدد معالم الفقرة .

النموذج ثنائي المعلم:

من الانسب البدء بالنموذج ثنائي المعلم لانه الاكثر تمثيلاً للنموذج الطبيعي، وللنموذج ثنائي المعلم الدالة:

$D\,ag\,(\theta - bg) = x$

اذ يمكن تحيد D اصطلاحاً، والشائع ان لها قيمة تساوي (1.7) وذلك لأن$P_g(\theta)$ للمنحنيات الطبيعية واللوغاريتمية لا تختلف عن بعضها البعض بأكثر من (0.01) لأي من قيم θ أي درجة السمة الكامنة (lord & Novick , 1968). والمعالم a_g و b_g لها الدور نفسه في النماذج اللوغاريتمية جميعها، وكذا الحال في نموذج المنحنى الطبيعي. وبالاسماء هي: معامل الصعوبة ومعامل التمييز، وبتعويض قيمة X في الصيغة العامة للنموذج اللوغاريتمي نصل الى صيغة النموذج ثنائي المعلم:

$$P_g(\theta) = \frac{e^{D_{ag}(\theta - bg)}}{1 + e^{D_{ag}(\theta - bg)}} \qquad (14\text{-}15)...............$$

النموذج احادي المعلم

يمكن تصوّر النموذج احادي المعلم على انه حالة خاصة للنموذج ثنائي المعلم الذي يكون لفقراته كلها معلم التمييز نفسه، ولان الفقرات كلها لها المعلم نفسه فإننا نرمز له بالثابت a لا المتغير a_g ،ودالة المنحنى المميز للفقرة احادية المعلم هي:

$$P_g(\theta) = \frac{e^{D_a(b-\theta)}}{1 + e^{D_a(b-\theta)}} \qquad (15\text{-}15)...............$$

ويمكن كتابتها بالصورة الآتية أيضاً:

$$P_g(\theta^*) = \frac{e^{(D^* - b_g^*)}}{1 + e^{(\theta^* - b_g^*)}} \qquad (16\text{-}15)...............$$

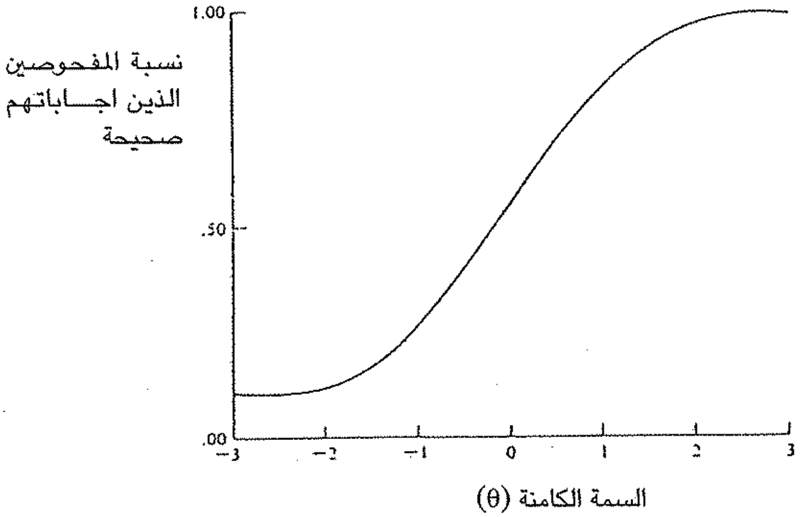
حيث ان $D a\,\theta = \theta$ و $D\,a\,bg = bg^*$. لذا فإن θ^* تعد تدريج جديد للقدرة θ،ويعبر عن bg^* على التدريج الجديد. وعند كتابة صيغة الدالة بهذه الطريقة فإن الصيغة تؤكد ان نسبة المفحوصين الذين يستجيبون بصورة صحيحة وعلى وفق النموذج احادي المعلم تعد دالة لقدرة المفحوص وصعوبة الفقرة. واشتغل راش بمفاهيم تختلف عن المنحنى المميز للفقرة، وطوّر نموذج سمّي باسمه وهو مكافئ للنموذج احادي المعلم. وقدم رايت(wright , 1968) تفسيراً لاتجاه راش التطوّري، ومن الممكن ملاحظة انه يوجد نموذج منحنى طبيعي مشابه للنموذج احادي المعلم.

النموذج ثلاثي المعلم:

قد تظهر مشكلة النموذج الطبيعي والنماذج اللوغاريتمية احادية المعلم وثنائية المعلم عند تطبيقها على بيانات فقرات الاختيار من متعدد أو الصح والخطأ، وذلك لان هذه الصيغ جميعاً تسمح بالاجابات الناتجة عن التخمين. والمشكلة في النماذج احادية وثنائية المعلم أن قيمة $pg(\theta)$ تقترب من الصفر كلما انخفضت قيمةθ ، ومع ذلك يمكننا توقع حتى للمفحوص ذو القدرة المنخفضة ان تكون النسبة التي تجيب بصورة صحيحة ستكون اكبر من الصفر وذلك لان بإمكان هؤلاء المفحوصين ان يخمنوا الاجابة الصحيحة، لذلك فإنه يمكن استخدام النموذج ثلاثي المعلم بسبب هذا العامل، ومعادلة هذا النموذج هي:

$$Pg(\theta) = Cg + \frac{(1-C_g)\, e^{Dag(\theta - bg)}}{1+ \, e^{Dag(\theta - bg)}} \qquad (17\text{-}15)\ldots\ldots\ldots$$

ويطلق على Cg اسم معلم التخمين، ويبين شكل (15 –7) مثال لهذا المنحنى، حيث Cg =0.1، ويلاحظ ان المنحنى المميز للفقرة يقترب من 0.1 بانخفاض القدرة، وعلى هذا فإنه سيجيب على الاقل 10% من المفحوصين على الفقرة اجابة صحيحة وحتى عند القيم المنخفضة جداً لـ θ.



شكل (15 – 7): النموذج اللوغاريتمي ثلاثي المعلم.

حساب المعالم

يجب حساب قيم المعالم bg، ag، cg، لكل فقرة اعتماداً على احد النماذج اللوغاريتمية، وعلى الأقل فإن هنالك طريقتين شائعتين للحساب هما دالة الأرجحية القصوى والهامشية أو التقريبية. وهنا سنقدم مراجعة لطريقة الأرجحية القصوى. وقدم سوامنياثان (swaminathan , 1983) عرض وافي وتقني لطرائق الارجحية القصوى، وقدم وود ووينغرسكي ولورد(lord , 1976& wood , wingersky) برنامج لوجست (logist) المحوسب المستخدم لحساب الارجحية القصوى للنموذج ثلاثي المعلم. ويحسب هذا البرنامج معالم الفقرة للفقرات جميعها ودرجات السمة الكامنة للمفحوصين جميعهم، ويشار الى هذه الطرائق بالارجحية القصوى المشتركة، ولهذه الطريقة شروط عدة يجب توافرها عند استخدامها مع النموذج اللوغاريتمي ثلاثي المعلم وهي:

أولاً: تتطلب عدد اساسي من المفحوصين للتقدير الدقيق، وفي دراسة (Hulin , lissak & drasgow,1982) التي بينت أن دقة الحساب باستخدام المنحنى المميز للفقرة ثلاثي المعلم لاختبارات مؤلفة من 30 فقرة و 60 فقرة، ووجدوا ان تقديرات المعالم تصبح أكثر دقة عند الانتقال من (200 الى 500 الى 1000) مفحوص. والدقة لا تتحسن بالدرجة نفسها عند الانتقال الى (2000) مفحوص، كذلك فإن دراسة هولين وزملاؤه تقترح تكبير حجم العينة للحصول على تقديرات دقيقة لـ ag، ونتائج مشابهة توصل اليها (lord , 1968) و (Ree & jenesen, 1980) و (swaminathan & Gifford, 1979) إذ استنتجوا الحاجة الى عينة كبيرة لحساب معالم الفقرة (وأشرنا في السابق الى استنتاج هولين بأنه يلزم حجم عينة اصغر بكثير عندما يكون الهدف الرئيسي تقدير θ) .

ثانياً: لغاية الآن لا نعرف ان حساب المعالم يكون متسقاً، فالطرائق التي تؤدي الى تقديرات متسقة لمعالم الفقرة لا نعلم ما إذا كانت تقديراتها تقترب اكثر واكثر من القيم الحقيقية للمعالم بزيادة حجم العينة اكثر واكثر.

وطريقة اخرى لحساب معالم النموذج ثلاثي المعلم هي الارجحية القصوى الهامشية (وفي الادب المبكر كان يشار اليها باسم الارجحية غير المشروطة) وحتى الوقت الحالي فإن هذه الطريقة لها محددات عملية تعوق استخدامها، ومع ذلك فإن كل من (Bock & atkin ,1981) طورا طريقة لحساب معالم تتغلب على المشكلات العملية واستخدموها مع النموذج الطبيعي، وبينوا امكانية تطبيقه على النموذج ثلاثي المعلم. ويخدم برنامج (BILOG) المحوسب النماذج الاحادية والثنائية والثلاثية المعالم، ويوزع من قبل شركة Seientific soft ware والفائدة

الاساسية للارجحية القصوى الهامشية انها توفر حسابات متسقة للمعالم Kiefer).
wolfowitz, 1956)

والبرنامج الاكثر انتشاراً في امريكا الشمالية لحساب معالم النموذج احادي المعلم هو برنامج بيكال (BICAL) والذي طوره (Wright, Mead & Bell,1979)، ويحسب هذا البرنامج تقديرات الارجحية القصوى المشتركة (ويذكر كل من رايت وميد وبيل ان هذه التقديرات غير مشروطة ولكنهم يستخدمون هذا المصطلح بطريقة تختلف عن طريقة بوك واتكين، وهذه الطريقة لا توفر معالم متسقة، لذا فإن رايت ودوغلاس أوجدوا عامل تصحيح يجعل هذه التقديرات متسقة. وتستخدم الارجحية القصوى المشروطة في غرب اوروبا في حساب المعالم وحتى الوقت الحالي فإن هذه الطريقة تعد عملية فقط مع الاختبارات القصيرة التي فقراتها أقل من 30 أو 40 (Wainer , morgan & Gustafsson ,1980). والتطورات الحديثة جعلت من الممكن استخدام هذه الطريقة مع اختبارات مؤلفة من 80 إلى 100 فقرة (Gustafsson , 1980) وتكمن أهمية طريقة الارجحية القصوى المشروطة بأن تقديراتها للمعالم تكون متسقة.

تفسير مخرجات برنامج بيكال:

هذا الجزء مقتبس من مخرجات برنامج بيكال، وقد طور هذا البرنامج من قبل رايت وميد وبيل (wright , Mead & Bell , 1979) للتحليل وفق النموذج احادي المعلم (نموذج راش)، وقد اختير عرض مخرجات هذا البرنامج لانه الاسهل تفسيراً للمبتدئين والاكثر انتشاراً، وتتفق تفسيراته مع التفسيرات المقترحة من قبل معدي البرنامج، ويجب التذكير بأن البرنامج يقدم معلومات اخرى مهمة غير تلك المعروضة هنا.

وفي عرضنا هذا سنركز على ثلاثة انواع من المخرجات:

1. تقدير صعوبة الفقرة.
2. تقدير قدرة المفحوص.
3. معلومات حول مطابقة الفقرة.

ومثالنا يعتمد على استجابات 143 مفحوص على (20) فقرة اختبارية من نوع الاختيار من متعدد في الرياضيات (وللتنويه نذكر ان حجم العينة المستخدم للتوضيح أقل بقليل من أقل حجم لهذا البرنامج وهو 200 مفحوص) وقبل عرض الاقتباس المذكور سنصف باختصار طريقة الارجحية القصوى غير المشروطة (وسميت بالطريقة غير المشروطة من قبل رايت وميد وبيل ولكنها في الحقيقة طريقة ارجحية قصوى مرتبطة) ،وتستخدم هذه الطريقة مجموعة

معادلات تكون فيها صعوبة الفقرة ودرجة السمة الكامنة غير معروفة، ويبدأ تنفيذ الطريقة بحساب قيم أولية لتقديرات الصعوبة ودرجة السمة الكامنة، وتعد هذه القيم تخمينات حول تقديرات الارجحية القصوى غير المشروطة. ويستخدم البرنامج المحوسب هذه الطريقة بطريقة تؤدي الى مجموعة ثانية لتقديرات الصعوبة والسمة الكامنة، وتستخدم المجموعة الثانية لانتاج مجموعة ثالثة وهكذا، ويستمر تكرار الحساب حتى دورة اخرى يكون فيها التغير في تقديرات المعالم أقل ما يمكن، وتتضمن المجموعة الاخيرة تقدرات الارجحية القصوى غير المشروطة، ويبين جدول (15-6) في الاعمدة الخمس الاولى بيانات حول تقديرات صعوبة الفقرة، إذ يحتوي العمود الأول على أرقام الفقرات، وهي هنا من (1 إلى 20)، والعمود الثاني يعيد ترقيم الفقرات بطريقة جديدة مناسبة ،وفي هذا المثال يستخدم رقم الفقرة، ويقترح استخدام محتوى الفقرات كذلك، والعمود الثالث صعوبة الفقرة ويتألف من تقدير مستوى القدرة لكل فقرة على تدريج السمة الكامنة عندما يجيب 50%

جدول (15-6): بيانات صعوبة الفقرة من برنامج النموذج أحادي المعلم بيكال

الرقم المتسلسل	اسم الفقرة	صعوبة الفقرة	الخطأ المعياري	اخر تغير في الفرق
1	0001	2.799 -	.469	2.211-
2	0002	1.375-	0.291	1.086-
3	0003	0.566	0.193	0.447
4	0004	0.218-	0.217	0.172-
5	0005	0.172-	0.215	0.136-
6	0006	1.651-	0.317	1.304
7	0007	0.465-	0.228	0.368-
8	0008	0.037	0.210	0.029-
9	0009	0.314-	0.221	0.248-
10	0010	0.175	0.203	0.139
11	0011	0.037-	0.210	0.029-
12	0012	0.750	0.191	0.592
13	0013	0.414	0.196	0.327
14	0014	0.750	0.191	0.592
15	0015	0.092	0.205	0.073
16	0016	0.266-	0.219	0.210-
17	0017	1.682	0.192	1.328
18	0018	1.179	0.188	0.931
19	0019	0.007	0.208	0.006
20	0020	1.719	0.193	1.357

من المفحوصين بصورة صحيحة على الفقرة. بكلمات أخرى تعد هذه القيم تقديرات لـ bx لكل فقرة. ومن هذا العمود نرى أن الفقرة الأولى هي الأسهل إذ يجيب 50%من المفحوصين على الفقرة الذين درجة سمتهم الكامنة (-2.79)، بينما الفقرة (20) الأخيرة هي الأصعب إذ يجب إن تكون درجة السمة الكامنة (1.719)حتى يجيب 50% من المفحوصين على الفقرة بصورة صحيحة. والعمود الرابع يبين الخطأ المعياري لتقدير صعوبة كل فقرة. فللفقرة الأولى نكون على ثقة 68% بأن معلم الصعوبة يقع بين (-3.268) (وهي = -2.799 - 469 = 33.2) و =(-2.33) وهي (-2.799 + 5.469 = 0). العمود الأخير المسمى آخر تغير في الفرق يشير إلى اتجاه وحجم التعديل الذي أجري على تقدير الصعوبة بعد التكرار الاخير للبرنامج. ويرى أن قيم معاملات الصعوبة ثابتة نسبياً بعد التكرار الاخير، ولكن هذا التقدير للفقرة يتغير بصورة ملحوظة. ولبيانات صعوبة الفقرة أهمية خاصة عندما يريد مطور الاختبار بناء اختيار مميز ضمن مدى قدرة معين. وتفيد مثل هذه البيانات في بناء اختبارات الاختيار من متعدد وذلك للتأكد من أن درجة الصعوبة نفسها للفقرات جميعها.

ويبين الجدول (15–7) المعلومات المستخدمة لتحويل الدرجات الخام الى درجات قدرة على تدريج السمة الكامنة (القدرة الرياضية)، العمود الأول المسمى العد يمثل عدد المفحوصين الذين حصلوا على درجة خام معينة، والاخطاء المعيارية هي الاخطاء المعيارية لكل درجة قدرة. لاحظ ان هذه القيم هي الأقل في منطقة وسط التوزيع والاكبر عند القيم التقديرية المتطرفة. وتفيد بيانات درجات القدرة في تحديد درجات القطع ومعادلة درجات الاختبارات المختلفة (وستعرض هذه العمليات في الفصول التالية) .

ويبين جدول (15–8) البيانات التي تستخدم لتقييم مطابقة الاستجابات الفعلية للنموذج اللوغاريتمي احادي المعلم. وتفيد هذه البيانات معد الاختبار في تحديد الفقرات التي لا تطابق النموذج احادي المعلم (انظر: حسن المطابقة في الجزء المتعلق باختيار النموذج). وانظر أولاً الى جدول (15 –8 أ) اذ تم تقسيم المفحوصين الى ستة مجموعات، المجموعة الاولى لها أقل درجات والمجموعة السادسة لها أعلى درجات. ويبين الجدول نسبة المفحوصين الذين اجاباتهم صحيحة على كل فقرة. على سبيل المثال:

جدول (15 -7): تقديرات درجة السمة الكامنة المستخرجة من برنامج بيكال

جدول الدرجات المكافئة الكامل			
الدرجة الخام	العد	تقدير القدرة	الاخطاء المعيارية
19	18	3.48	1.12
18	14	2.60	0.81
17	11	2.05	0.68
16	15	1.64	0.61
15	9	1.30	0.56
14	16	1.00	0.53
13	12	0.73	0.51
12	12	0.48	0.50
11	2	0.24	0.49
10	6	0.0	0.49
9	3	0.24-	0.49
8	13	0.48-	0.50
7	6	0.73-	0.51
6	5	1.00-	0.53
5	1	1.30-	0.56
4	0	1.64-	0.61
3	0	2.05-	0.68
2	0	2.60-	0.81
1	0	3.48-	1.12

للفقرة 11 نرى ان 20%من المجموعة الأولى اجابوا عنها اجابة صحيحة، و 61% من المجموعة الثانية اجابوا عنها اجابة صحيحة، وهكذا ،فكلما ازدادت قدرة المجموعة نلاحظ زيادة في نسبة الذين يجيبون اجابة صحيحة على الفقرة (كما هو الحال في الفقرة 11) .

وهذه النسب الملاحظة للاستجابات الصحيحة قورنت بالنسب النظرية التي يجب ان تظهر فيما لو طابقت البيانات النموذج احادي المعلم. ويبين جدول (15–8 ب) النسب النظرية والفعلية لكل مجموعة، وتشير القيمة (15 ,0) الى درجة المجموعة الاولى على الفقرة 11 ان

النسبة الملاحظة للاستجابات الصحيحة اقل من 15%، وبكلمات اخرى فإن نسبة الاستجابات الصحيحة اقل من 15% . وبكلمات اخرى فإن نسبة الاستجابات الصحيحة من المنحى المميز للفقرة المتوقع لمجموعة هذه الدرجات يجب ان تكون 20 % + 15% = 35% وتشير النسبة المئوية للمجموعة الثانية على الفقرة 11 أنها اكبر ب 4% من النسبة المتوقعة، لذا فإن 61% من المجموعة اجابوا بصورة صحيحة على الفقرة و 57% منهم يجيبون بصورة صحيحة حسب ما هو متوقع من المنجنى المميز للفقرة. ويبين جدول (15– 8 جـ) المعامل الاحصائي للتوافق بين المنحنى المميز للفقرة الملاحظ والمتوقع في العمود المسمى " المطابقة بين " fit between)، ولتفسير هذا الاحصائي يجب العودة الى المتوسط صفر والانحراف المعياري واحد. فإذا تمعنا في الفقرات التي لها قيمة احصائي مطابقة 2.00 أو اكثر نجدها للفقرات (1، 4 ،7، 9، 12) إذ يبدو ان لهذه الفقرات قيم مطابقة مشكوك فيها (لاحظ انه لا يوجد اتساق في نسب المفحوصين الذين يجيبون بصورة صحيحة في المجموعات الستة، ان قيمة احصائي المطابقة يتأثر بحجم البعد عن القيمة المتوقعة للمنحنى المميز للفقرة، والتي تكون اوزانها مختلفة عبر مناطق مختلفة من المنحنى). واحصائي مطابقة اخر يظهره المنحنى يسمى اختبار ت – الكلي، وقد وصف مؤلفو هذا البرنامج المعامل على انه الاتفاق العام بين السمة المقاسة بالفقرة وتلك المقاسة بالاختبار ككل وثانية يجب ارجاع هذه القيم الى المتوسط صفر والانحراف المعياري واحد. وفي احصائي المطابقة هذا نجد أن للفقرات. 11و 17 قيم تزيد عن 2.00، مع ملاحظة انه للفقرة 11 معامل ارتباط بوينت بايسيريال بين الفقرة والدرجة الكلية هي الأعلى لفقرات الاختبار جميعها(0.62)، وسبب عدم المطابقة لمثل هذه الفقرة هو تمييزها الشاذ بالنسبة لفقرات الاختبار الاخرى، لذا فإنها لا تطابق النموذج أحادي المعلم الذي يتطلب تساوي تمييز الفقرات جميعها، لذا فإن عدم مطابقة هذه الفقرة لا يعزى إلى خلل في الفقرة بل العلاقة بالسمة الكامنة لا يتم تمثيلها بشكل مناسب بالنموذج احادي المعلم، واقترح مطوروا هذا البرنامج مراجعة الفقرات اذا كان احصائي المطابقة بين اكبر من 3 أو 4 أو مطابقة كلية أكبر من 2، وذلك إما بسبب اخطاء مفتاحية أو اخطاء طباعة أو أي عد م اتساق قد يسبب عدم المطابقة، وإن كان من غير الممكن خفض احصائي المطابقة لاسباب تقنية يجب تدقيق محتوى الفقرة للتأكد من ان الفقرة ضمن حقل المتغير المقصود، ويجب تذكير القارئ بأن هذه المقترحات مبنية على التجربة العلمية، ويمكن استخدام محكات اخرى للحكم على درجة المطابقة.

وما يثير الدهشة هنا هو ان معالم الفقرات تتصف بخاصية عدم التباين عند حسابها باستخدام احد نماذج السمة الكامنة، وفي أدب القياس يوجد تفسيرين على الأقل لعدم التباين، التفسير الأول قدمه هامبلتون وكوك (Hambelton & Cook , 1977) ولورد (Lord , 1980)والذي يعالج منحنى الفقرة على أنه منحنى انحدار، إذ اشار لورد الى انه وفي السياقات الاحصائية لا تتغير دالة الانحدارعندما يكون توزيع متغير الانحدار (وهو السمة الكامنة هنا) متغير. ومع ان هذا صحيح الا انه لا يعني ان عدم تباين المنحنى المميز للفقرة لا يتباين آلياً بالنسبة لاختيار المفحوصين

وقد عزا لورد (1980) ولورد ونوفيك (1968) عدم تباين هذا الى احادية البعد فقد وافقا على انه إذا تغير المنحنى المميز للفقرة عبر المجموعات الفرعية من المفحوصين مثل الذكور والاناث فإنه يوجد تغير منتظم على الفقرة g بين الذكور والاناث الذين لهم درجة السمة الكامنة نفسها لذا فإنه لا يوجد سبب لاعتبار الاختبار احادي البعد. لاحظ أن منظور عدم التباين في الاختبار ليس من خصائص نماذج السمة الكامنة ولكنه متطلب منطقي لاحادية البعد، ومن هذا المنظور فإن عدم التباين الاختيار يعد افتراض منتهك عن قصد في التطبيق. ونتيجة ذلك فإن التطبيقات التي تعتمد على افتراض عدم التباين يجب ان تتحقق من صحة هذا الافتراض.

حبك الاختبار

عند استخدام الاختبارات المقننة تظهر مشكلة عدم فائدتها لبعض المفحوصين فمثلاً يكون العديد من فقرات الاختبار المقنن صعبة جداً للمفحوصين ذوي القدرة المتدنية، ويلجأ العديد من هؤلاء المفحوصين الى التخمين في الاجابة على العديد من الفقرات، لذا فإن هذه الفقرات تتضمن اخطاء قياس اكثر من المعلومات القيمة عن المفحوصين، ايضاً افترض ان جامعة ما تقبل فقط 10 % من مجموع المتقدمين لامتحان القبول، وفي مثل هذا الموقف لا تساعد الفقرات السهلة في التمييز بين المفحوصين الذين تقع درجاتهم قريباً من درجة القطع بالنسبة للاختيار، إذ ان معظم المفحوصين يميلون للاجابة عن مثل هذه الفقرات إجابة صحيحة. وما يقترح من مثل هذه الأمثلة هو المزاوجه بين صعوبات الفقرات وقدرات المفحوصين بدلاً من الحصول على اجابة كل مفحوص على كل فقرة. وهذا هو هدف الحبك (الاختبارات المكيفة) وتقنياً فإن استراتيجيات حبك الاختبار تعتمد على مفهوم يعرف ب " المعلومات " وما تبقى من هذا الجزء يركز على هذا المفهوم، وللوصول الى فهم افضل لهذا المفهوم افترض الأمثلة

الآتية: يريد مطور الاختبار بناء اختبار عدد فقراته خمس بهدف تحديد مواقع المفحوصين على تدريج السمة الكامنة الذي يمتد من – ∞ الى + ∞، مع هذا فإن مطور الاختبار يعلم ان درجات السمة الكامنة للمفحوصين في عينته تمتد من – 3.00 الى + 3.00، وتتوزع اعتدالياً حول نقطة الصفر، افترض أنه من الممكن بناء فقرات لها دالة خطوة للمنحنى المميز للفقرة، وتتغير هذه النقطة عندما تظهر الخطوة لكل منحنى مميز للفقرة. في المحاولة يعد مطور الاختبار خمسة فقرات لكل منها منحنى مميز للفقرة للخطوة عند θ = صفر. وعند تطبيق الاختبار نجد ان 50 % من المفحوصين كانت درجاتهم الكامنة أقل من صفر وأن 50% من المفحوصين كانت درجاتهم الكامنة أكبر أو تساوي صفر من الواضح ان هذا الاختبار يزودنا بمعلومات حول ما إذا كانت السمة الكامنة للمفحوص أقل من صفر أو لا وعلى ذلك فإن جميع معلومات الاختبار تهتم بنقطة صفر التدريج.

جدول (15-9): نتائج اختبار مؤلف من خمس فقرات على تدريج جوتمان

الدرجة	نمط الاستجابة	موقع درجة القدرة
صفر	– – – – –	$1.0- > \theta \geq 3-.0$
1	– – – – +	$.5- > \theta \geq 1.00$
2	– – – + +	$0.0 > \theta \geq 0.5-$
3	– – + + +	$0.5 > \theta \geq 0.0$
4	– + + + +	$1.0 > \theta \geq 0.5$
5	+ + + + +	$3.0 > \theta \geq 1.0$

وفي المحاولة الثانية في بناء الاختبار تظهر خطوات المنحنى المميز للفقرة عند النقاط - 1.0 , 0 , 0.5 , 0.10. والآن تتراوح الدرجات الخام بين صفر و 0.5 ، ويبين جدول (15-9) نماذج الاستجابة ومواقع المفحوصين الذين حصلوا على كل درجة لاحظ أن الاختبار يزودنا بمعلومات عن مواقع المفحوصين الذين سمتهم الكامنة أقل من – 1.0 أو أعلى من 1.0، ومع ذلك فإنه يزودنا بمعلومات جيدة عن مواقع المفحوصين بين – 1.0 و + 1.0.

وفي المحاولة الثالثة بنيت الفقرات بخطوات -2، -1، 0، +1، +2 وثانية، فإنه لوحظت

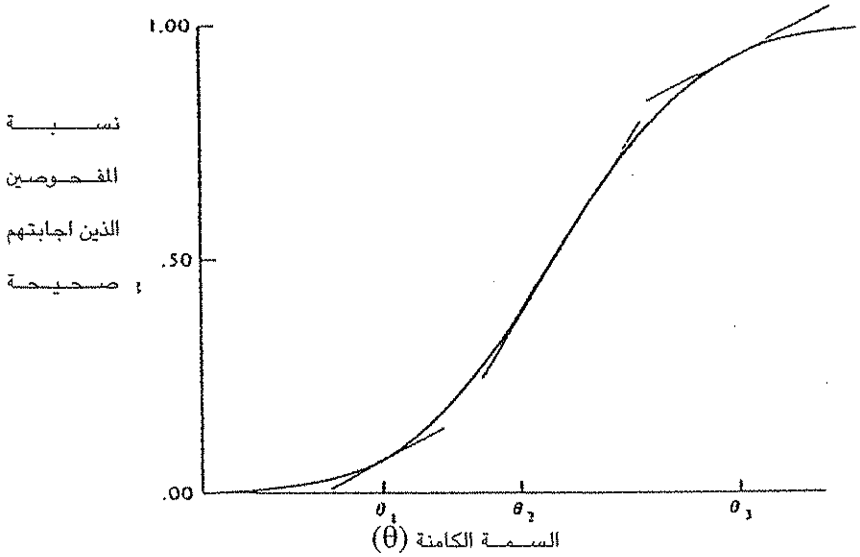
احتمالات الدرجة الملاحظة جميعها من صفر الى خمسة، ويبين جدول (15–10) نماذج الاستجابة والمواقع المناظرة لكل درجة ملاحظة، ومن الواضح ان هذا الاختبار يقدم معلومات اكثر من الاختبار الثاني لتحديد مواقع المفحوصين في المنطقة $1.0 \ge \theta \ge -3$ و $3.0 \ge \theta \ge 1.0$ ، كذلك فإنه يزودنا بمعلومات أقل عن تحديد مواقع المفحوصين في المنطقة $1 \ge \theta \ge -1$، ويكتشف مطور الاختبار الآن المعضلة، فمن اجل بناء اختبار مفيد في الخطوة الأولى فإنه يجب تحديد المواقع على تدريج السمة الكامنة التي تميز بشكل جيد عبر نقاط التدريج والخطوة التالية هي بناء اختبار يزودنا بمعلومات تحدد مواقع المفحوصين في هذه المناطق

من الواضح من الأمثلة السابقة ان الاختبار يزودنا بمعلومات تفاضلية عبر المناطق المختلفة على تدريج السمة الكامنة وتمييز افضل يكون عبر المفحوصين بدرجات سمة كامنة في المناطق التي يقدم فيها الاختبار معلومات اكبر عبر المناطق المختلفة، وعلاوة على ذلك فإن المعلومات المستقاة من الاختبار تعتمد على المنحنى المميز للفقرة، ولهذا السبب فإنه يمكننا الحديث عن معلومات الفقرة .

ومع ذلك فان العرض اعلاه يغطي المعنى العام للمعلومات، وهذا المفهوم له تعريف رياضي دقيق لنوع $P'_g(\theta)$ تمثل المشتقة الأولى للمنحنى المميز للفقرة. وللقارئ الذي ليس له معرفة

جدول (15–10) نتائج اختبار مؤلف من خمس فقرات على تدرج جوتمان

الدرجة	نمط الاستجابة	موقع درجة القدرة
صفر	- - - - -	2.0 - > θ ≥ 3-
1	- - - - +	1.0 - > θ ≥ 2.0-
2	- - - + +	0.0 > θ ≥ 1.0-
3	- - + + +	1.0 > θ ≥ 0.0
4	- + + + +	2.0 > θ ≥ 1.0
5	+ + + + +	3.0 > θ ≥ 2.0



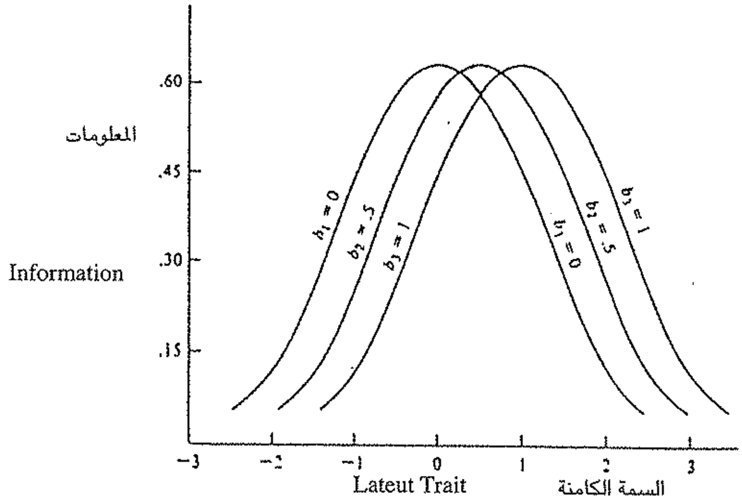
شكل (15-8): توضيح لخطوط المماس على المنحنى المميز للفقرة

بالمعادلات التفاضلية بكفية معرفة ان المشتقة الاولى عند أي قيمة θ تساوي ميل الخط المرسوم عند مماس المنحنى المميز للفقرة عند النقطة $P_g(\theta)$. ويبين الشكل(15–8) مثل هذه المماسات، ويتبين من الشكل ان ميل المنحنى عند θ_1و θ_3 قليل نسبياً، وكذلك قيم المشتقة عند θ_1 و θ_3، بينما ميل المماس عند θ_2 كبير نسبياً، وكذلك قيمة المشتقة. لندع Q_g والمعلومات التي تزودنا بها الفقرة عند أي نقطة g على تدريج السمة الكامنة هي:

$$I_g(\theta) = \frac{[P'_g(\theta)]^2}{[P_g(\theta)]\,Q_g(\theta)} \qquad \text{(8-15)}$$

وبسط المعادلة هو المشتقة الاولى لـ $P_g(\theta)$، لذا فان المعلومات تكون اكبر عندما تكون قيمة المشتقة اكبر، ونتيجة ذلك فإن الفقرة في الشكل (15 –8) تزودنا بمعلومات عند θ_2 اكبر من تلك عند θ_1 و θ_3 ، وذلك لان المشتقة الاولى عند θ_2 اكبر ويمكن حساب المعلومات عند كل نقطة على تدريج السمة الكامنة، وذلك للحصول على المعلومات التي تزودنا بها الفقرة عند كل قيمة من قيم θ .وعند رسم منحنى $I_g(\theta)$ مع θ نحصل على دالة معلومات الفقرة، ويمثل الشكل (15 –9) دوال معلومات الفقرة للفقرات الثلاث للمنحنى المميز للفقرة، للنموذج ثنائي المعلم. فالفقرات الثلاثة لها قيمة a_g نفسها بينما تتغير قيم bg لهاواقل قيمة صعوبة للفقرات الثلاث b_1 =صفر، ومنحنى معلومات هذه الفقرة هو الابعد الى جهة اليسار، مع ملاحظة أنه

يصل الى اعلى قيمة عند 0 = صفر، وهذا يعني ان هذه الفقرة تزودنا بمعلومات اكثر عن المفحوصين الذين درجة السمة الكامنة لهم تساوي صفر. لاحظ أيضاً ان منحنى معلومات الفقرة الأولى يقع دون 0.15 لمن درجة السمة الكامنة لهم اقل من -2.0 وأعلى من 2.0، وهذا يعني ان الفقرة تزودنا بمعلومات قليلة عن المفحوصين الذين تقع سمتهم الكامنة في هذه المنطقة .



شكل (15 -9): توضيح لمنحنى المعلومات لفقرات ثلاث، b_1= صفر ، و b_2 = 0 و 1 = b_3

وعند الحصول على المعلومات المحددة يمكننا وصف حبك الاختبار بصورة مفصلة، فاحد اساليب الحبك يتطلب مصرف اسئلة كبير معاير على تدريج مشترك ويطبق على المفحوصين بوساطة الحاسوب، وتبدأ الطريقة بتقدم كل مفحوص الى عدة فقرات، وتستخدم استجاباتهم في تكوين تقدير أولي لدرجة السمة الكامنة للمفحوص، وبعدها يقدم الحاسوب الفقرة التي تعطينا اكبر قدر من المعلومات عند درجة السمة الكامنة والمساوية لقدرة المفحوص الكامنة المحسوبة في البداية، واعتماداً على درجة السمة الكامنة المستنبطة من الفقرة تقدم له فقرة جديدة، وأحد قواعد انتهاء الاختبار بهذه الطريقة هو تحديد محك دقيق لتقدير السمة الكامنة، ويستمر الاختبار حتى نحصل على تقدير دقيق للسمة الكامنة.

المعلومات التي تقدمها طرائق الاختبار والتصحيح المختلفة

اثبت مفهوم المعلومات اهميته عند:

1. مقارنة صيغ التصحيح المختلفة للاختبار نفسه
2. مقارنة اختبارات مختلفة تقيس السمة نفسها

3. مقارنة طرائق الاختبار المختلفة مثل الاختبارات المحبوكة والمقننة التقليدية. وعند اجراء مثل هذه المقارنات فاننا نستخدم دوال معلومات الاختبار ومعلومات التصحيح، لذا سنبدأ هذا الحزء بتعريف هذه الدوال.

دالة معلومات الاختبار هي مجموع دوال معلومات الفقرة لفقرات الاختبار جميعها، وبالرموز هي:

$$\sum_{g} I_g(\theta) = t(\theta) \quad (19\text{-}15)..............$$

طريقة تصحيح معينة والمعلومات التي يزودنا بها، لذا فمن المناسب التمييز بين المعلومات التي تزودنا بها طريقة تصحيح معينة والمعلومات التي يزود،نا الاختبار، وسنرمز الى معلومات التصحيح بالرمز $I(\theta, x)$، وذلك لاي صيغة طريقة تصحيح معينة والمعلومات التي يزودنا مستخدمة، ويمكننا ان نبين أن $I(\theta, x) \le I(\theta)$، اي الحد الاعلى للمعلومات التي نحصل عليها من التصحيح. إن ابسط صيغة للتصحيح هي اعطاء درجة واحدة لكل اجابة صحيحة وصفر لكل اجابة خاطئة، ويمكن التعبير عن صيغة التصحيح هذه على النحو:

$$\sum U_g = X \quad (20\text{-}15)..............$$

وترمز U_g الى درجة الفقرة g التي قد تكون صفر أو واحد. وقد بيّن بيرنباوم (**Birnbaum,** 1968) ان صيغة التصحيح هذه للنموذج احادي المعلم تؤدي الى اقصى قدر من المعلومات يمكن الحصول عليها وينسحب هذا على الاختبارات التي لفقراتها جميعاً الصيغة نفسها للمنحنى المميز للفقرة بغض النظر عن صيغته. وللنموذج ثنائي المعلم تكون صيغة التصحيح $\sum a_g u_g = x$, وهذه ينتج عنها اكبر دالة معلومات ويمكننا ان نبين ولاي نموذج انه لا يمكننا الحصول على $I(\theta)$ لقيم θ جميعها بأي صيغة تصحيح وذلك بسبب:

1. الجمع الموزون لدرجات الفقرات

2. واستخدام الاوزان نفسها للمفحوصين جميعهم. فمثلاً استخدام الاوزان الآتية مع النموذج اللوغاريتمي ثلاثي المعلم:

$$W_g = \frac{D_{a_g}[P_g(\theta) - C_g]}{[I - C_g]\, P_g(\theta)} \quad (21\text{-}15)..............$$

ينتج عنه منحنى معلومات مساوياً لمنحنى معلومات الاختبار لقيم θ جميعها، وهذه الاوزان تعتمد على $p_g(\theta)$، لذا فان هذه الاوزان تختلف باختلاف المفحوصين.

والسؤال المهم يكمن في فقدان المعلومات عندما لا يستخدم نظام الاوزان المثالي كصيغة معتمدة في التصحيح. ففي حالة النموذج ثنائي المعلم على سبيل المثال السؤال المهم هو نقص المعلومات عند استخدام الصيغة $x = u_g$ للتصحيح بدلاً من الصيغة $\sum a_g u_g = y$ ويمكن اختبار هذا السؤال في الفقرة التالية من خلال مفهوم الفعالية النسبية .

افترض ان x و y تمثلان قياسين مختلفين للسمة نفسها، فقد تكون x و y طريقتين مختلفتين في تصحيح الاختبار نفسه أو انها درجات مشتقة من اختبارات مختلفة تقيس السمة الكامنة نفسها. وتتحدد دالة الفعالية النسبية للقياسين بـ:

$$\frac{I(\theta, x)}{I(\theta, y)} = RE(\theta,x,y) \qquad (22\text{-}15)................$$

وتحسب دالة الفعالية النسبية كما في دالة المعلومات عند كل نقطة على تدريج السمة الكامنة وتشير الى اي القياسين يعد الافضل لقياس السمة الكامنة عند كل نقطة على التدريج، ولأي درجة سمة كامنة عندما تكون $1< RE(\theta,x,y)$ فإن X يكون قياس أفضل، وبصورة مشابهة عندما تكون درجة السمة الكامنة $1 > RE(\theta,x,y)$ فإن Y يكون قياس أفضل للسمة الكامنة.

وتستخدم الفعالية النسبية لاختبار فقدان المعلومات عندما لا تستخدم صيغة التصحيح نظام أوزان مثالي للفقرات. فعلى سبيل المثال يطرح سؤال حول فقدان المعلومات عند استخدام صيغة التصحيح $\sum ug = y$ بدلاً من الصيغة $\sum a_g u_g = y$، وهنا تتحدد الفعالية النسبية بالصيغة $\frac{I(\theta, x)}{I(\theta, y)}$ واشار بيربناوم الى هذه الصيغة الخاصة لدالة الفعالية النسبية على انها دالة الفعالية. وحيث ان الدالة هي$I(\theta) \geq I(\theta, X)$ فإن قيمة دالة الفعالية تكون أقل أو تساوي 1. ووجد هامبلتون وتراوب **(Hambelton & Tranb , 1971)** ان دالة الفعالية غير الموزونة للنموذج ثنائي المعلم تكون اكبر من 0.85، ويمكن تفسيرها لتعني ان طول الاختبار المصحح بدون صيغة أوزان يجب زيادته بنسبة $\frac{1}{0.85}$اي (1.18) من اجل الحصول على المعلومات التي نحصل عليها من التصحيح الموزون المثالي. بشكل اخر فإن استخدام صيغة التصحيح غير الموزونة تكافئ اختزال طول الاختبار بنسبة 100 (1−85.0%) (أي 15%) عند استخدام الاوزان المثالية. وبين هامبلتون وتراوب ان استخدام الصيغة $\sum u_g = x_g$ مع النموذج ثلاثي المعلم ينتج عنه دالة فعالية تتراوح قيمته بين 0.60 الى 70.0 عند مستويات القدرة المنخفضة.

وتستخدم الفعالية النسبية ايضاً لمقارنة الاختبارات المقننة التقليدية بالاختبارات المكيفة (المحبوكة). وكما هو ملاحظ فإن الاختبارات المكيفة هي احدى الطرائق المتبعة في المزاوجة بين صعوبة الفقرات ومستوى قدرة المفحوصين، ويفضل هذا الاسلوب الاختباري لان الفقرات الصعبة جداً والسهلة جداً تزودنا. بمعلومات قليلة عن مستوى القدرة الكامنة للمفحوصين. وقد وصف لورد (lord , 1980) ثلاثة انواع مختلفة من الاختبارات المكيفة. احدها الاختبار مرن المستوى الذي يتألف من فقرات ترتب وفق صعوبتها ويبدأ المفحوص بالاجابة عن فقرات متوسطة الصعوبة، وعند الاجابة عنها اجابة صحيحة يطرح على المفحوص فقرات اكثر صعوبة، وإلا فقرات اقل صعوبة عند الاجابة الخاطئة عليها. ويبين لورد (lord , 1970) كيفية ترتيب الفقرات في كتيب الاختبار إذ تجمع الفقرات وفق تقديمها للمفحوص في الاختبار مرن المستوى. ويمكن اجراء مقارنة افتراضية بين اختبار مرن المستوى وآخر تقليدي وذلك بتحديد معالم المنحنى المميز لكل فقرة. وذلك بتخصيص أو تحديد قانون التصحيح وحساب الفعالية النسبية، وستبين النتائج ان تدريج السمة الكامنة للاختبار مرن المستوى تكون معلوماته اكبر من معلومات الاختبار التقليدي في بعض مناطق تدريج السمة الكامنة، وهذا يساعد في التأشير على الاهداف التقييمية التي يتميز بها الاختبار مرن المستوى على الاختبار التقليدي. وبالتغيير المنتظم لمعالم الفقرة يمكننا تحديد القانون تجريبياً لتصميم اختبار مرن المستوى أكثر فاعلية وتتألف هذه القوانين التجريبية من قيم مناسبة محددة لمعالم الفقرة. وإذا افترضنا صلاحية مصرف الاسئلة فمن الممكن بناء اختبارات تكوينية مرنة ومن المهم محاولة تطبيق مثل هذه الاختبارات من اجل فحص طرائق تطبيق الاختبارات مرنة المستوى مع مفحوصين حقيقيين.

ويمكن استخدام الطرائق المشابهة لهذه والموصوفة في الاختبارات مرنة المستوى في تطوير قواعد تصميم الاختبارات ثنائية المرحلة والاختبارات المحبوكة. يتألف الاختبار ثنائي المرحلة من اختبار تمهيدي قصير وعدة اختبارات اطول (اختبارات المرحلة الثانية). إذ تستخدم درجة المفحوص. على الاختبار التمهيدي في تحديد اختبار المستوى الثاني الذي سيتقدم اليه المفحوص وتقترح النتائج التي عرضها لورد (1980) انه بالامكان بناء اختبارات ثنائية المرحلة واخرى محبوكة تزودنا بمعلومات اكثر عند مستويات القدرة المختلفة مما تزودنا به الاختبارات التقليدية.

وأخيراً من الضروري تمييز ان نظرية الاستجابة للفقرة والمفاهيم المرتبطة بالمعلومات والفعالية النسبية يمكن تطبيقها في مقارنة اختبارات لم تبنى بالاساس وفق نماذج السمة الكامنة فعلى سبيل المثال لخص لورد دراسة تقارن بين سبعة اختبارات لفظية للصف

السادس باستخدام الفعالية النسبية، وبينت النتائج انه في الاختبارات المتتابعة للتقدم التربوي (step) السلسلة الثانية والمستوى الرابع للصيغة – أ – اي الاختبار الفرعي في القراءة مقارنة باختبار ميتروبوليتان التحصيلي (MAT) للمستوى المتوسط – الصيغة – و –. ان اختبارSTEP يوفر معلومات اكثر عن الفئة متدنية التحصيل بنسبة 20 % من اختبار MAT وللطلبة من الفئات الاخرى فان اختبارMAT يقدم معلومات اكثر من اختبار STEP وتشير مثل هذه الدراسات الى قدرة نظرية الاستجابة للفقرة في صنع قرارات تتعلق بملائمة الاختبارات بغض النظر عن الطريقة الاساسية المعتمدة في تطوير الاختبار .

الخلاصة:

يقدم هذا الفصل مقدمة لنظرية الاستجابة للفقرة. وفي البداية تم عرض مفاهيم اساسية عدة، والمنحنيات المميزة للفقرة والسمات الكامنة والاستقلال المحلي والابعاد والاختبار المتحرر من عينة القياس. وبعدها تم عرض النموذج اللوغاريتمي الطبيعي ومعالمه –صعوبة الفقرة b_g وتمييزها a_g. وكذلك تم توضيح العلاقة بين معالم نظرية القياس التقليدية ومعالم النموذج اللوغاريتمي الطبيعي، وكذلك العلاقة بين درجات السمة الكامنة والدرجات الحقيقية.

وبعدها عرضت نماذج لوغاريتمية ثلاثة. نموذج المعلم الواحد (أو راش) وهو النموذج الابسط، ويحسب بوساطته معلم صعوبة الفقرة فقط، ويفترض ان الفقرات جميعها تميز بالتساوي وان المفحوصين من ذوي القدرة المتدنية لا يجيبوا اجابات صحيحة بالتخمين. ويشبه النموذج ثنائي المعلم النموذج اللوغاريتمي الطبيعي ويتضمن معلم للتمييز ويتضمن النموذج ثلاثي المعلم معلم شبه الصدفة بالاضافة الى معلمي الصعوبة والتمييز، لذا فإنه يعد النموذج الاقل تقييداً. ونوقش أيضاً حساب المعالم بشكل مختصر اتبع بعرض تفصيلي لمخرجات النموذج المحوسب بيكال: وهو برنامج محوسب يستخدم للحسابات المرتبطة بالنموذج احادي المعلم.

وتضمن الفصل وصف تطبيقات عدة لنظرية الاستجابة للفقرة، هي مصارف الاسئلة والاختبارات المحبوكة ومقارنة دقة قياس ادوات مختلفة عند نقاط مختلفة على تدريج السمة الكامنة. كذلك تم تقديم مفاهيم عدم تباين الاختبار والمعلومات والفعالية النسبية.

تمارين :

1/ اعتماداً على الشكل (15-3) بين ان الفقرتين الاولى والرابعة مستقلة محلياً للمفحوصين الذين درجة السمة الكامنة لهم θ_2 . وبيّن أن لفقرتين الثانية والثالثة مستقلة وبين أن احصائياًعند أي درجة سمة كامنة في المدى $\theta_2 < \theta < \theta_3$.

2/ افترض ان بيانات اختبار طابقت النموذج اللوغاريتمي ثنائي المعلم. لاول فقرتين (θ) $P_1 = 0.5$, $P_2(\theta) = 0.7$ لدرجة سمة كامنة معينة، بالاضافة لهاتين الفقرتين $= 0.45$ $P(+, +1/\theta)$ ، $P(-, +1/\theta) = 0.05$ و $P(-, -1/\theta) = 0.25$ فهل تقترح مثل هذه النتائج احادية البعد ؟ لماذا نعم؟ ولماذا لا؟

3/ فيما يأتي انماط استجابة افتراضية على خمس فقرات:

نمط الاستجابة	الفقرة				
	1	2	3	4	5
أ	+	+	-	-	-
ب	-	-	-	-	-
ج	+	+	+	+	+
د	+	-	-	-	-
هـ	+	+	+	+	-
و	+	+	+	-	-

هل تنسجم انماط الاستجابة هذه مع الادعاء بان هذه الفقرات احادية البعد وان المنحنيات المميزة للفقرة هي دوال خطوة وان كان كذلك فارسم المنحنيات المميزة للفقرات الخمس .

4/ تبين النتائج التالية معامل بايسيريال للفقرة – السمة الكامنة P_g، وصعوبة الفقرة P_g لاربع فقرات، بافتراض التوزيع الطبيعي للقدرة θ لاربع فقرات، بافتراض لتوزيع الطبيعي للقدرة θ بمتوسط صفر وانحراف معياري 1. احسب a_g و b_g لكل فقرة. وعلى اساس هذه النتائج هل P_g تعاني من قصور كمقياس لصعوبة الفقرة؟

الفقرة	P_g	P_g
1	0.8	0.7
2	0.6	0.7
3	0.5	0.7
4	0.4	0.7

5/ تتضمن الجداول التالية معلومات متوقعة لمخرجات برنامج بيكال لخمس عشر فقرة، و 474 مفحوص طبقت عليهم الفقرات لمعايرة كل من صعوبات الفقرات ومعالم قدرة المفحوص للنموذج احادي المعلم.

الرقم المتسلسل	اسم الفقرة	المجموعة 1	المجموعة 2	المجموعة 3	المجموعة 4	المجموعة 5	المجموعة 6
المنحني المميز للفقرة							
1	0001	0.35	0.55	0.64	0.82	0.83	0.91
2	0002	0.84	0.86	0.94	0.99	1.00	0.98
3	0003	0.06	0.18	0.08	0.11	0.18	0.54
4	0004	0.49	0.54	0.47	0.56	0.61	0.71
5	0005	0.86	0.89	0.98	0.99	0.99	1.00
6	0006	0.51	0.71	0.89	0.86	0.92	0.99
7	0007	0.21	0.28	0.53	0.72	0.85	0.91
8	0008	0.56	0.86	0.88	0.97	1.00	0.99
9	0009	0.12	0.09	0.16	0.11	0.24	0.33
10	0010	0.26	0.35	0.28	0.35	0.51	0.69
11	0011	0.23	0.43	0.44	0.47	0.59	0.82
12	0012	0.40	0.68	0.78	0.88	0.94	0.96
13	0013	0.82	0.94	0.92	0.91	1.00	0.99
14	0014	0.25	0.17	0.36	0.49	0.48	0.68
15	0015	0.18	0.46	0.65	0.75	0.87	0.93

مدى الدرجات	7-1	8-8	9-9	10-10	11-11	14-12
متوسط القدرة	-0.50	0.19	0.55	0.92	1.33	2.05
المتوسط الزائي	0.1	0.1	0.2	0.3	0.6	0.6
الانحراف المعياري (بالدرجات الزائية)	2.1	1.5	0.9	1.4	1.4	1.3
عدد افراد المجموعة	77	65	85	80	71	96

رقم الفقرة	اسم الفقرة	المجموعة 1	المجموعة 2	المجموعة 3	المجموعة 4	المجموعة 5	المجموعة 6
الانحراف عن المنحنى المميز للفقرة المتوقع							
1	0001	0.05-	0.01-	0.01-	0.10	0.03	0.02
2	0002	0.02	0.04-	0.01	0.04	0.03	0.00-
3	0003	0.01	0.08	0.06-	0.09-	0.09-	0.11
4	0004	0.22	0.13	0.03-	0.03-	0.08-	0.10-
5	0005	0.00-	0.03-	0.03	0.02	0.01	0.01
6	0006	0.08-	0.03-	0.09	0.01	0.02	0.05
7	0007	0.09-	0.17-	0.01-	0.10	0.13	0.07
8	0008	0.13-	0.04	0.02	0.07	0.07	0.02
9	0009	0.07	0.00	0.04	0.06-	0.00-	0.06-
10	0010	0.10	0.09	0.05-	0.08-	0.02-	0.00-
11	0011	0.01	0.08	0.00-	0.6-	0.04-	0.05
12	0012	0.11-	0.00	0.03	0.06	0.08	0.03
13	0013	0.02	0.05	0.00-	0.03-	0.04	0.01
14	0014	0.09	0.10	0.02	0.06	0.05-	0.01-
15	0015	0.17-	0.05-	0.05	0.06	0.11	0.6

+ = أكثر

- =أقل

احصائي مطابقة الفقرات							
الرقم المتسلسل	الخطأ التزايد بسبب المطابقة	المطابقة بين	اختبارات كلي	WED متوسط المربع	متوسط المربعات الانحراف المعياري	فهرس التميز	معامل بوينت بايسيريال
1	0.0	0.33	1.88-	0.92	0.04	1.17	0.42
2	0.0	0.65	1.00-	0.86	0.14	1.01	0.24
3	0.0	2.78	1.07-	0.93	0.07	1.14	0.33
4	0.09	4.47	5.10	1.18	0.03	0.28	0.15
5	0.0	0.35	1.03-	0.83	0.17	1.14	0.24
6	0.0	1.64	2.18-	0.85	0.07	1.28	0.42
7	0.0	3.40	4.45-	0.85	0.04	1.52	0.54
8	0.0	2.85	2.39-	0.79	0.10	1.46	0.45
9	0.02	1.76	0.63	1.04	0.07	0.67	0.20
10	0.03	1.82	1.55	1.06	0.04	0.76	0.29
11	0.00	0.17	0.17	1.01	0.03	1.00	0.37
12	0.0	1.69	2.37-	0.86	0.06	1.35	0.44
13	0.0	0.64	0.84-	0.89	0.13	0.94	0.21
14	0.0	1.34	1.03	1.04	0.04	0.86	0.31
15	0.0	3.09	4.21-	0.84	0.04	1.52	0.53

أ- حدد ثلاثة فقرات تبين انماط استجابة المفحوصين أن مطابقتها ضعيفة للنموذج ؟ والى ماذا يعود غياب المطابقة.

ب- على متصل واحد، وباستخدام البيانات في الجدول لمجموعات القدرة الست ارسم المنحنى المميز للفقرات 1، 2، 3، 4، على التوالي، ثم ارسم النقاط الستة لكل فقرة مبيناً الخصائص الحقيقية للمفحوصين عند مستويات القدرة هذه الذين اجابوا أجابة صحيحة باستخدام المتصل نفسه. وحدد لكل فقرة منطقة القدرة التي ينحرف الأداء عليها بشكل واضح عن الاداء المتوقع .

ج- رتب الفقرات 1، 3، 4، بدلالة صعوباتها اعتماداً على ملاحظاتك لمواقعها على المتصل المرسوم في النقطة ب.

د- حصل المفحوص أ على درجة خام 10 في الاختبار، وحصل المفحوص ب على 11 نقطة ما تقدير الدرجة الحقيقية المحسوبة لكل منهما

6/ يبين الجدول ادناه المعلومات [$Ig(\theta)$] التي زودنا بها اختبار مؤلف من 9 فقرات عند النقاط التسعة على تدريج السمة الكامنة ولكل فقرة منها a_g =0.5، بينما قيم b_g تتراوح بين – 2.0 الى + 2.0 . وقد حسبت اشكال المعلومات لدالة النموذج اللوغاريتمي ثنائي المعلم، وقد تم الحصول على نتائج مشابهة جداً بوساطة النموذج الطبيعي استخدم اشكال المعلومات لحساب كل من:

أ) معلومات الاختبار التي يزودنا بها الاختبار المؤلف من الفقرات التسع احسب قيم معلومات الاختبار لقيم السمة الكامنة المبينة في الجدول

ب) معلومات الاختبار التي يزودنا بها الاختبار المتضمن الفقرات التسع. عند a_g= 0.5و b_g = صفر .

b_g									θ
2.0	1.5	1.0	0.5	0.0	0.5-	1.0-	1.5-	2.0-	
0.02	0.03	0.05	0.07	0.09	0.12	0.15	0.17	0.18	2.0-
0.03	0.05	0.07	0.09	0.12	0.15	0.17	0.18	0.17	1.5-
0.05	0.07	0.09	0.12	0.15	0.17	0.18	0.17	0.15	1.0-
0.07	0.09	0.12	0.15	0.17	0.18	0.17	0.15	0.12	0.5-
0.09	0.12	0.15	0.17	0.18	0.17	0.15	0.12	0.09	0.0
0.12	0.15	0.17	0.18	0.17	0.15	0.12	0.09	0.07	0.5
0.15	0.17	0.18	0.17	0.15	0.12	0.09	0.07	0.05	1.0
0.17	0.18	0.17	0.15	0.12	0.09	0.07	0.05	0.03	1.5
0.18	0.17	0.15	0.12	0.09	0.07	0.05	0.03	0.02	2.0

7/ يبين الجدول ادناه معلومات الاختبار لاربعة اختبارات عند تسع نقاط على تدريج السمة الكامنة. في الاختبارين الاول والثاني كانت قيمة ag للفقر ات جميعها = 1.0، بينما كانت قيمتها في الاختبارين الثالث والرابع = 1.5 ولكل زوج من الاختبارات الاولى قيم b_g لها هي مثل قيم b_g للفقرات التسع في الفرع الاول (أ) من السؤال السادس، بينما الثاني يتألف من تسع فقرات لكل منها b_g =صفر. وبناءً على هذه

النتائج واجابتك على سؤال 6، وبافتراض عدم الاهتمام بدقة قياس المفحوصين عند $\theta < -2.0$ أو $\theta > 2.0$ فماذا تستنتج عن مستخدمية تضمين فقرات بمدى صعوبة أوسع في الاختبار؟ وفي اجابتك على هذا السؤال افترض ان الفقرات التي لها $a_g > 1.0$ نادرة وان الفقرة التي لها $a_g = 1.0$ لها تقريباً معامل ارتباط بايسيريال مع السمة الكامنة = 8.0 وللفقرات التي لها $a_g = 0.5$ و 1.5 لها معاملات ارتباط بايسيريال مناظرة 0.45، 0.83.

1.5=a_g		1.0 =a_g		θ
bg = صفر	bg المختلطة	bg صفر	bg المختلطة	
0.36	3.37	0.81	2.06	-2
1.26	4.48	1.71	2.67	-1.5
3.96	4.92	3.42	3.04	-1.0
9.99	5.06	5.49	3.21	-0.5
14.67	5.09	6.48	3.26	0.0
9.99	5.06	5.49	3.21	0.5
3.96	4.92	3.42	3.04	1.0
1.26	4.48	1.71	2.67	1.5
0.36	3.37	0.81	2.06	2.0

8/ تتألف الصيغة الاختبارية المختصرة في البرامج الاختبارية من فقرات تجريبية طبقت جنباً إلى جنب مع الصيغة المنتظمة للاختبار. وقد اختيرت فقرات الصيغة المنتظمة من ملف فقراته مدرجة على تدريج مشترك وطابقت بياناتها نموذج السمة الكامنة اللوغاريتمي ثنائي المعلم لكلا الاختبارين المنتظم والتجريبي، واختير التدريج بحيث يكون متوسط السمة الكامنة صفر وانحرافها المعياري 1. كيف يمكن استخدام هذه النتائج في تحديد مواقع الفقرات التجريبية المحسوبة معالمها على التدريج نفسه التي تقع عليه معالم الفقرات المنتظمة.

الفصل السادس عشر

16

الكشف عن تحيز الفقرة

الفصل السادس عشر

الكشف عن تحيز الفقرة

من المسلم به أن درجات الاختبار تتأثر بمصادر التباين إضافة إلى البناء الذي يقيسه الاختبار. وإن لم يكن هذا حقيقياً، فإن درجات الاختبار ستكون صادقة وثابتة تماماً، ولأن مصادر التباين التي ليست ذات علاقة لا يمكن إلغاؤها، فمن المهم أن لا تكون متحيزة ضد المجتمعات الفرعية من المفحوصين، مثل الذكور مقابل الإناث. ومثل هذا التحيز يكون موجوداً إذا كان ضمن مجموعة من المفحوصين الذين لهم نفس الموقع على متصل السمة التي يقيسها الاختبار، كذلك فإن توزيع مصادر التباين غير المناسبة يميز أو يفاضل بين المجموعتين الفرعيتين. ولفهم كيفية ظهور مثل هذا، خذ بعين الاعتبار المواقف الآتية:

1- قد توجد فقرة في اختبار ذكاء مشهور تطرح هذا السؤال: "ما الذي ستفعله إذا عثرت على محفظة جيب تعود لأحد الأشخاص في مخزن الأمتعة؟". الإجابة الصحيحة هي الإخبار عن ذلك إلى الموظف المسؤول عن ذلك في المخزن. يقال أن مثل هذه الفقرة قد تكون متحيزة ضد الأطفال ذات الدخل المتدني، إذ أن أخذ المال إلى الوالدين قد تبدو إجابة أو ردة فعل ملحوظة عند مثل هؤلاء الأطفال أكثر من غيرهم.

1. قد تطلب فقرة في اختبار مفهوم الذات من الطلبة الإجابة على تدريج ليكرت على العبارة: "أن بشرتي لها مظهر جميل، وثانية قد يقال عن مثل هذه الفقرة أنها متحيزة إذ أن لها معنى مختلف عن المراهقين البيض أكثر منها للسود، لذا فإن الإجابات المتشابهة قد لا تؤثر على مواقف متشابهة على سمة الاهتمام.

2. توجد فقرة في اختبار التحصيل الرياضي تقول "لدى رالف ربعي دولار، وباكيت العلكة يكلفه 0.25 دولار، فكم علبة يمكن شراؤها؟" فإن كانت هذه الفقرة تطبق على أطفال لا يستطيعون قراءة الإنجليزية ضمن مجموعة الأطفال الكلية، فإن مثل هذه الفقرة تقيس مهارات مختلفة لأطفال خاضعين لموقف اختباري واحد.

إن هنالك فكرتين أساسيتين من هذه النقاشات حول تحيز الفقرة:

الأولى: هي أن أداء الأفراد الخاضعين للاختبار على الفقرة قد يتأثروا بمصادر تباين غير تلك المتعلقة بالبناء المقيس.

الثانية: هو أن هذه المصادر الإضافية للتباين تؤثر على الأداء بطريقة مغايرة وبشكل منتظم لبعض المجموعات الفرعية من المجموعة الكلية التي خضعت للاختبار. ومن الممكن أيضاً تأثر درجات المجتمعات الفرعية بمصادر التباين المختلفة، علاوة على أن المدى الذي تكون فيه الفقرة متحيزة ضد مجتمعات فرعية خاصة، وقد لا يكون التحيز واضحاً من خلال عملية فحص محتوى الفقرة. فعلى سبيل المثال: قد يكون للأفراد المفحوصين من مجتمع فرعي معين في بعض المواقف أداء أفضل على الفقرات التي تم الحكم عليها من قبل مراجعي المحتوى وصنفت على أنها فقرة متحيزة ضد هذه الفئة من المفحوصين أكثر من الفقرات التي صنفت على أنها غير متحيزة. لذا فإن مطور الاختبار أو مستخدمه الذي يرغب ببحث واكتشاف الفقرات المتحيزة ضد مجموعات معينة يجب عليه إجراء دراسة تجريبية على الفقرات التي يعتقد الباحث تحيزها.

وهنالك هدفان للدراسات التي تبحث في تحيز الفقرات:

الأول: بحث ما إذا كانت درجات الاختبار تتأثر بمصادر التباين المختلفة في المجتمعات الفرعية كافة. وإذا حُكم على درجات الاختبار للمجتمعات الفرعية كافة أنها تتأثر بمصادر التباين نفسها، فإن الهدف الثاني هو تحديد ما إذا كانت أي من هذه المصادر غير ذات العلاقة تعطي أي ميزة متحيزة لبعض المجتمعات الفرعية. ومن خلال هذين الهدفين يمكننا صياغة تعريف للفقرات غير المتحيزة على أنها:

1. تتأثر بمصادر التباين نفسها في كلا المجتمعين الفرعيين.
2. أن الأفراد من المستوى نفسه على البناء المقيس يكون توزيع مصادر التباين غير ذات العلاقة هو نفسه لكلا المجموعتين الفرعيتين.

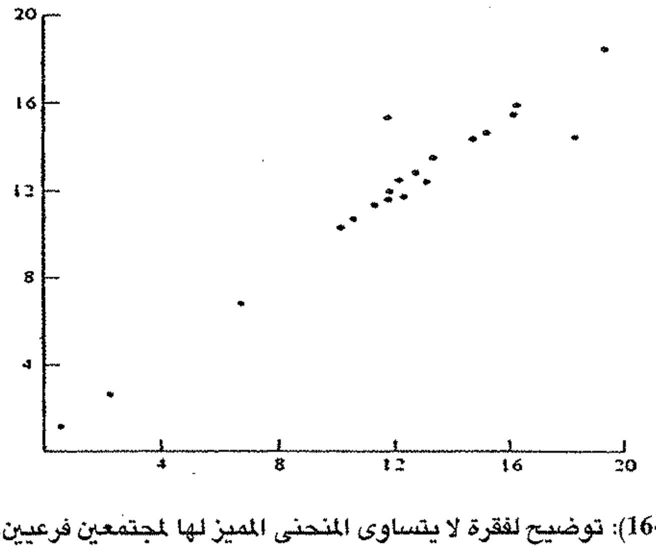
ويهتم هذا الفصل بطرائق بحث واكتشاف تحيز الفقرات. لن نحاول مناقشة الطرائق جميعها، ولكننا سنختار أكثرها شيوعاً وواعدة في مجال دراسات تحيز الفقرات. وقد قدم جنس (Jensen,1981) عرضاً شاملاً لهذه الطرائق. وسنناقش ثلاثة فئات من هذه الطرائق: نظرية السمة الكامنة ومربع كاي وصعوبة الفقرة. ولم تجري أي محاولة لتقديم التفاصيل الإجرائية الكاملة لهذه الطرائق، إلا أنه تم ذكر مصادر للمهتمين بتفاصيل هذه الإجراءات. وعند عرض أي من هذه الطرائق سنجري مقارنة بين مجتمعين فرعيين ولكن هذه الطرائق تستخدم للمقارنة بين أكثر من مجموعتين.

الطرائق المعتمدة على نظرية الاستجابة للفقرة:

عند استخدام نظرية الاستجابة للفقرة للتحقق من تحيز الفقرات، فإنه يتم الحكم على

الفقرات بأنها غير متحيزة إذا كانت المنحنيات المميزة للفقرات (ICCs) هو نفسه لكلا المجتمعين الفرعيين. ومن ثم فإن صعوبة الفقرات تكون متساوية لأفراد المجتمعين الفرعيين. وبهذا فإن مصادر التباين غير ذات العلاقة تؤثر على المجتمعين الفرعيين بالطريقة نفسها. إضافة إلى ذلك، بما أن المنحنيات المميزة للفقرة هي نفسها لكلا المجتمعين الفرعيين، فإن الفقرات تقيس السمة نفسها لكلا المجتمعين الفرعيين، وهكذا، فإن كانت مجموعة الفقرات غير متحيزة على وفق تعريف نظرية السمة الكامنة فستكون غير متحيزة وفقاً للتعريف المقدم آنفاً. (وبالفعل فإن هذا الاستنتاج يكون صحيحاً فقط إذا رغبنا الافتراض بأن المفحوصين المتجانسين في السمة الكامنة هم أيضاً متجانسين في البناء الذي نهتم به، فإن كان الأفراد المتجانسين على السمة الكامنة غير متجانسين على البناء الذي نهتم به، فإن الاختبار ليس له صدق بناء أساسي. واستخدام الاختبار المشكوك في صدق بنائه يماثل إجراء دراسة ثبات لاختبارات مشكوك في صدقها. فإن كلا الفعاليتين لا قيمة لهما). والفقرات المتحيزة وفقاً لتعريف نظرية السمة الكامنة قد لا تكون متحيزة وفقاً لتعريفنا هذا. وسنناقش هذه القضية بتفصيل أكثر في نهاية هذا الفصل. ويوضح الشكل (16-1) فقرة لا يتساوى المنحنى المميز لها لمجتمعين فرعيين.

ولأن المنحنيات المميزة لمجموعة فقرات غير متحيزة لا تتباين عبر المجموعات الفرعية فإن معظم المعاملات التي تعتمد على نظرية السمة الكامنة تعد قياساً لمدى تباين المنحنيات المميزة للفقرات عبر هذه المجموعات. ويتطلب حساب هذه المعاملات خطوتين: الأولى: تقدير معالم الفقرة لكل مجتمع فرعي والتعبير عنها على التدريج نفسه، والثانية حساب معامل تحيز الفقرة للفقرات جميعها.



شكل (16-1): توضيح لفقرة لا يتساوى المنحنى المميز لها لمجتمعين فرعيين.

تدريج معامل الفقرة:

قبل حساب معامل تحيز الفقرة يجب حساب معالم الفقرة على التدريج نفسه لكل مجتمع فرعي. احدى الطرق المناسبة للحساب هو تقييد تقديرات (b_g) بحيث يكون متوسطها (صفر) وانحرافها المعياري (واحد) (LORD,1980). وهذا يعرف باسم معايرة قيم (B_g's)، وآلياً يتم تحديد مواقع المعالم المحسوبة على التدريج نفسه لكلا المجموعتين، وتصلح هذه الطريقة للنماذج اللوغاريتمية الثلاثة ولنموذج المنحنى الطبيعي (وفي حالة النموذج أحادي المعلم من الضروري فقط التقيد بمستوى صعوبة = صفر فقط).

وطريقة بديلة تستخدم المعايرة على قيم θ، أي على درجات السمة الكامنة، وفي هذه الطريقة يتم تدريج قيم السمة الكامنة المحسوبة بشكل منفصل لكل مجموعة فرعية. ولكل مجموعة فرعية يكون متوسط السمة الكامنة صفر وانحرافها المعياري واحد. وبما أن تدريج السمة الكامنة مختلف لكل مجموعة عن الأخرى وكذلك تدريجات الصعوبة والتمييز، إلا أنه من الممكن تحويل قيم الصعوبة والتمييز للمجموعة الفرعية (1) لتصبح على تدريج المجموعة الفرعية الثانية، وذلك باستخدام المعادلتين:

$$\hat{b}_{ig} = K\,\hat{b}_{ig} + m \qquad \text{(16-1أ)} \; ....$$

$$\hat{a}_{ig} = \frac{a_{ig}}{k} \qquad \text{(16-1ب)} \; ....$$

وفي كلا المعادلتين تمثل $\hat{a}_{1g}$ و $\hat{b}_{1g}$ قيم المجموعة الفرعية الناتجة عن معايرة θ، وقيم θ و k و M هي ثوابت تحسب على النحو المبين أدناه، وقيم $\hat{a}_{1g}$ و $\hat{b}_{1g}$ تكون على تدريج $\hat{a}_{2g}$ و $\hat{b}_{2g}$ نفسه، أي الحسابات التي حصلنا عليها للمجموعة 2 والناتجة عن معايرة θ. ويمكن حساب قيمة K بحساب ميل المنحنى البياني لقيم $\hat{b}_g$ مقابل b^*_g . ويحسب المنحنى والقاطع لهذا الخط البياني كما يأتي:

$$\hat{k} = ({}^2\sigma_1 - {}^2\sigma_2) + \sqrt{({}^2\hat{\sigma}_1 - {}^2\hat{\sigma}_2) + 4\,P_{12}\,{}^2\hat{\sigma}_1\,{}^2\hat{\sigma}_2} \qquad \text{(16-2أ)} \; ....$$

$$\text{و} \quad \hat{m} = \overline{b_2} - \overline{b_1}k \qquad \text{(16-1ب)} \; ....$$

وفي المعادلة (16-2أ) تمثل $\hat{\sigma}^2_1$ التباين عبر الفقرات $b_1{}^*g$، بينما تمثل $\sigma_2{}^2$ التباين عبر الفقرات $b_2{}^*g$ ويمثل P_{12} الارتباط عبر الفقرات $b_1{}^*g$، و $b_2{}^*g$، وفي المعادلة (16-2ب) تمثل b_1، و b_2 المتوسطات الفقرات $\hat{b}_{2g}$ و $\hat{b}_1{}^*g$، على التوالي. (وقد أشار إيرونسون-(Iron-

(son,1982أن هذا الإجراء لحساب k و m غير مقنع تماماً، وقد ناقش بحثاً كاملاً لتحسينها). وهذه الطريقة تعمل مع النماذج احادية المعلم والثنائيـة وكذلك ثلاثية المعلم. فللنموذج أحادي المعـلم $m = 1$ ويجب ملاحظة أن المعلم c_g معلم شبه الصدفة أو التخمين لا يتأثر بتغير التدريج ولا داعي لمعادلة قيم c_g لكلا المجتمعين الفرعيين.

مقارنة المنحنيات المميزة للفقرات:

ما إن يتم التعبير عن قيم المعالم على التدريج نفسه، فإنه تجري مباشرة مقارنة المنحنيات المميزة للفقرات لكلا المجموعتين. وهنالك طرائق عدة للمقارنة فيما بينها. وباستخدام النموذج اللوغاريتمي ثلاثي المعلم اقترح لورد (Lord, 1980) اختبار الفرضية الصفرية بأن $a_{2g} = a_{1g}$ و $b_{2g} = b_{1g}$. ويجب ملاحظة أن هذا الاختبار مرتبط بطريقة حساب ومعايرة b_g's لكلا المجموعتين. ولصعوبة تقدير c_g بدقة فإنها مستثناة من هذا الاختبار. وهذا الاختبار يتم إجراؤه للفقرات جميعها، وهذا يؤدي إلى إحصائي مربع كاي، والذي يستخدم في اختبار الفرضية الصفرية. وتعد كل فقرة يتم رفضها باختبار الفرضية الصفرية فقرة متميزة، وتعد قيم إحصائي مربع كاي مقياساً لدرجة تحيز الفقرة. وقد قام لورد (Lord, 1980) بتطوير اختبار باستخدام النموذج ثلاثي المعلم، ويمكن استخدام اختبار مشابه في حالة النموذج ثنائي المعلم. ويمكن اختبار الفرضية الصفرية للنموذج أحادي المعلم بـ:

$$Z = \frac{b_{2g} - b_{1g}}{\sqrt{\hat{\sigma}^2_{b1g} + \hat{\sigma}^2_{b2g}}}$$

حيث تمثل $\hat{\sigma}^2_{b1g}$ الخطأ المعياري لصعوبة الفقرة. للفقرة g وللمجموعة i والقيمة الحرجة لـ $\mp Z_{\alpha/2}$ هي المئين $100(1-\alpha / 2)$ للتوزيع المعياري الطبيعي.

واقترح روندر (Runder,1977) حساب المساحة بين (ICC's) للمجموعتين الفرعيتين كمقياس للفرق بين منحيي(ICC's) الاثنين، ويمكن تقريب هذه المساحة باستخدام المعادلة:

$$A_g = \sum_{\theta=-4.00}^{\theta=4.00} 0.005 \left| P_{2g}(\theta) - P_{1g}(\theta) \right| \qquad (3\text{-}16)\ldots\ldots\ldots\ldots$$

وفي المعادلة (3-16) تشير كل من $P_{1g}(\theta)$ و $P_{2g}(\theta)$ إلى قيمة (ICC's) لكلا المجموعتين الفرعيتين. ويمكن حساب قيمة $Pi_g(\theta)$ لكل قيمة θ من (4.00– إلى + 4.00) في خطوات تباينها0.005 لكل مرة باستخدام المعادلة (3-16):

$$A_g \text{(باشارتها)} = \sum_{\theta=-4.00}^{\theta=4.00} 0.005 \, [P_{2g}(\theta) - P_{1g}(\theta)] \qquad (3\text{-}16)$$

والتي تؤدي إلى قياس مساحة بإشارة (تحمل إشارة)، إلى أن الفرق العددي بين القياس الذي يحمل إشارة والذي لا يحمل إشارة يظهر عندما يتقاطع المنحنيان المميزان للفقرتين كما في شكل (16-1). ولأن إشارة القيمة المطلقة في المعادلة (16-3) تحوّل القيم السالبة لـ $P_{2g}(\theta) - P_{1g}(\theta)$ إلى قيم موجبة، فإن الفروق الناتجة في المعادلة (16-3) تكون جميعها موجبة6 بينما المعادلة (16-4) لا تحوّل القيم السالبة إلى قيم موجبة لذا فإن الفروق الموجبة والسالبة يتم موازنتها إلى حدٍ ما، وقيم المساحات المقاسة التي لا تحمل إشارة تكون أكبر من تلك التي تحمل إشارة. وكنتيجة فإن الفقرات التي لها نفس (ICC's) والفقرات التي لها فروق موزونة تماماً وتحمل إشارة تؤدي إلى A_g (بإشارة) = صفر، ويبدو أن هذا غير مرغوب به حتى إذا تم موازنة الفروق فإن الفقرات تبقى متحيزة.

وكبديل لمقياس المساحة عند روندر اقترح لين وزملاؤه **(Linn, etal,1981)** إحصائي متوسط مربع تحيز الفقرة دالة لـ $[P_{2g}(\theta) - P_{1g}(\theta)]^2$ المسحوبة من θ= −3 إلى θ= 3 ويمكن استخدام كلا من الصيغتين الموزونة وغير الموزونة إلا أن الإحصائي الموزون يكون تقديره أكثر دقة. وتجد تفاصيل هذه الطريقة في(Linn, etal,1981)، وأبحاث أكثر حداثة للإحصائي الموزون تجدها في(Levine,1981) وفي (Wardrop& Levin& Linn,1982).

أي من معاملات تحيز الفقرة هو الأكثر ملائمة؟ فهل من الأفضل تدوين الإحصائي الذي طوره لورد لاختبار $b_{2g} = b_{1g}$ و $a_{2g} = a_{1g}$ أم قياس المساحة مثل تلك التي طورها روندر أو لين وزملاؤه؟ وعلى الرغم من أن محكاً آخر أثبت أنه مهماً في عملية الاختيار إلا أن(ICC's) لها دلالة معنوية مختلفة تماماً وتبقى بصورة جوهرية هي نفسها (Linn, etal,1981). فعلى سبيل المثال إذا كانت a_g =1، b_g = 3.5 و C = 0.2 في المجموعة الأولى، في حين كانت ag = 0.5، bg =5.0، و Cg= 0.2 في المجموعة الثانية، فإن قيمة $P_{2g}(\theta) - P_{1g}(\theta)$ لن تكون أكبر من 0.5 لاي قيمة (θ) بين (−3.0، و 3.0). وفي هذه الحالة الفرق الأكبر في المعلم قد يكون مؤشر مضلل للتحيز. ويبدو أن هذه الحقيقة تفضل أسلوب روندر ولين وزملاؤه على طريقة لورد.

مثال يستخدم نظرية الاستجابة للفقرة:

لتوضيح استخدام نظرية الاستجابة للفقرة في اختبار تحيز الفقرات تم تحليل استجابات

عينة تألفت من (300) مفحوص (150 ذكور و 150 إناثا) أجابوا على 10 فقرات. ويبين الجدول (16-1) تقديرات لمعالم الصعوبة والتمييز في نموذج المنحنى الطبيعي لعينتي الذكور والإناث. (وتم الحصول على هذه التقديرات باستخدام المعادلتين (15-8) و (15-9) بتعويض معامل ارتباط بونيت بايسيريال بين الدرجة الكلية ودرجة الفقرة لـ (P_g) وحساب نسبة الإجابة الصحيحة لصعوبة الفقرة. (وقد استخدم هذا الإجراء لأهداف توضيحية فقط، ويمكن الاستغناء عن استخدامه في دراسات تحيز الفقرة).

وقد تم الحصول على قيم $\hat{b}_{1g}^*$ و $\hat{a}_{1g}^*$ التي تصف عينة الذكور من خلال المعايرة على (θ) الذكور. أما قيم $\hat{b}_{2g}$ و $\hat{a}_{2g}$ التي تصف عينة الإناث فقد تم الحصول عليها من خلال المعايرة على (θ) الإناث. وبما أن التقديرات المعلمية للذكور والإناث ممثلة على تدريجات مختلفة، فلا يجوز مقارنتها مباشرة. ولكي نتمكن من مقارنتها يجب تحويل (a^*_{1g} و $\hat{b}_{1g}^*$) بتطبيق المعادلتين (16-1أ) و (16-1ب)، والذي يتطلب حساب قيم ($\hat{k}$) و ($\hat{m}$). ويبين الجدول (16-1) قيم ($\hat{b}_1$, $\hat{\sigma}_1^2$, b_2 , $\hat{\sigma}_2^2$, $\hat{P}_{12}$) الضرورية لحساب كل من k و m. وبالتعويض بالمعادلة (16-2أ) نحصل على:

$$\hat{K} = \frac{(1.343\text{-} 1.055) + \sqrt{(1.343 - 1.055) + 4\,(0.769)\,(1.055)^2\,(1.343)}}{2\,(0.769)\,(1.027)\,(1.158)}$$

$$= 0.798$$

وبالتعويض بالمعادلة (16–2 ب) نحصل على:

$$\hat{m} = 0.985 - (0.798)\,(\text{-}1.513) = 0.222$$

وبالتعويض لقيم $\hat{k}$ و $\hat{m}$ في المعادلتين (16-1أ) و (16-1ب) نحصل على:

$$\hat{b_{1g}} = 798 = b_1{}^*g + 0.222$$

$$\hat{a_{1g}} = \frac{\hat{a_1^*}g}{.798}$$

جدول (16-1): تقديرات الصعوبة والتمييز لعينات الذكور والإناث.

الفقرة	$\hat{b}_1^* g$	$\hat{a}_1^* g$	$\hat{b}_{2g}$	$\hat{a}_{2g}$
	الجنس			
	الذكور		الاناث	
1	3.726-	0.527	-1.467	1.198
2	0.115-	1.000	0.222	0.723
3	1.479-	2.528	1.602-	0.989
4	2.966-	0.424	1.734-	0.592
5	2.614-	0.571	3.202-	0.589
6	1.375-	1.730	1.278-	0.909
7	0.633-	0.881	0.139	0.718
8	1.196-	0.928	0.428-	0.963
9	0.943-	1.920	0.624-	0.903
10	0.083-	1.235	0.115	0.686

$1.513 - = \bar{b}_1$ $\quad$ $0.985- = \bar{b}_2$

$1.343 = \hat{\sigma}^2{}_2$ $\quad$ $1.055 = \hat{\sigma}^2{}_2$

$0.769 = \hat{\rho}_{21}$

وذلك لتحويل تدريج الذكور إلى تدريج الإناث

ويدون جدول (16–2) قيم $b_{1g}^{\hat{}}$ و $a_{1g}^{\hat{}}$ التي تم الحصول عليها من المعادلتين (16-1أ) و (16-1ب)، وكذلك ($b_{2g}^{\hat{}}$) و ($a_{2g}^{\hat{}}$) و a_g "أي مقياس المساحة لروندر" لتحيز الفقرة. وكانت مساحة التحيز الأكبر للفقرة (1) بينما كانت أقل مساحة للتحيز للفقرة 7. ولغاية الآن ونتيجة للتجربة لا يوجد دليل أو محك لتقرير متى تكون الفقرة متحيزة، ويبدو أن الإجراء العام هو فحص الفقرات ذات درجة التحيز الكبيرة نسبياً، ومحاولة إعادة كتابة هذه الفقرات لإزالة تحيزها، وبهكذا أسلوب نجد أن الفقرات (1و5) وربما (3) لها درجات تحيز كبيرة.

كيف ستحقق تقنيات السمة الكامنة هدفي بحوث تحيز الفقرة اللذين أشرنا إليهما سابقاً.
لقد أكدنا انه في حالة تساوى المنحنيات المميزة للفقرات للمجتمعات الفرعية فإن هذه الفقرات تكون غير متحيزة على وفق تعريف نظرية السمة الكامنة والتعريفات المفهومة ضمنياً في دراسات تحيز الفقرة. كذلك فإن الفقرات المتحيزة وفقاً لتعريف نظرية السمة الكامنة ليس شرطاً أن تكون متحيزة وفقاً لتعريف آخر للتحيز. وكما بين هنتر (Hunter,1975) أن تطابق بيانات الفقرات مع النموذج أحادي البعد في نظرية السمة الكامنة في حالة كون مجموعة الفقرات متعددة الأبعاد يؤدي إلى تغير قيم المعالم عبر المجموعات الفرعية المختلفة. لذلك فإنه عندما تتغير المنحنيات المميزة للفقرات للنموذج أحادي البعد (في نظرية السمة الكامنة) عبر المجموعات الفرعية، فقد يؤشر هذا إلى أن الفقرات متعددة الأبعاد. ومن وجهة نظرنا فإن تعدد الأبعاد لا يؤكد التحيز عندما تكون :

1. السمات الكامنة الأخرى لا علاقة لها بالبناء الذي يتم قياسه.
2. توزيع هذه السمات مختلف بين المجموعات الفرعية.

لذا فإن تباين المنحنيات المميزة للفقرات لا يؤكد بالضرورة تحيز الفقرات.

جدول (2-16): قياس المساحة (Ag) لتحيز الفقرة.

الفقرة	الجنس			
	الذكور		الاناث	
	$\hat{b}_1^* g$	$\hat{a}_1^* g$	$\hat{b}_{2g}$	$\hat{A}_{2g}$
1	2.753-	.672	1.467-	1.198
2	0.130	1.252	0.222	0.723
3	0.959-	3.167	1.602-	0.989
4	2.146-	0.531	1.734-	0.592
5	1.865-	0.715	3.202-	0.598
6	0.875-	2.167	1.278-	0.909
7	0.283-	1.103	0.139	0.718
8	0.733-	1.162	0.428-	0.963
9	0.531-	2.405	0.624-	0.903
10	0.155	1.546	0.115	0.686

أما السبب الثاني لتباين المنحنيات المميزة للفقرات فقد يكون لقياس سمات كامنة أحادية البعد مختلفة لكلا المجتمعين. ومن وجهة نظرنا لا تعد هذه الحالة مؤشراً لتحيز الفقرات فكيف إذن يمكن التحقق من سبب تباين المنحنيات المميزة للفقرات؟ كمبدأ، يمكن التحقق من السبب من خلال تحديد البعد الذي تقيسه الفقرات في كلا المجتمعين الفرعيين. فإن كانت الفقرات أحادية البعد في كل مجموعة، فإن المنحنيات المميزة للفقرات المختلفة تكون ناتجة عن قياس سمات كامنة مختلفة، وإن كانت الفقرات متعددة الأبعاد في كل مجموعة فإن تباين المنحنيات المميزة للفقرات يكون ناتجاً عن تعددية الأبعاد. لذا فإنه لا توجد إجراءات مناسبة كلية لتأسيس البُعدية لذلك، فعندما تكون المنحنيات المميزة للفقرات مختلفة لمجتمعين فرعيين يكون التفسير ضبابي (غير واضح) نوعاً ما.

وهناك نقطة أخرى مهمة يجب التركيز عليها تتعلق بكون الفقرات أحادية البعد في كل مجموعة على حدة ولكنها متباينة عبر المجموعات المختلفة. والنتيجة السابقة تؤكد كون السمة الكامنة المقيسة تختلف من مجموعة لأخرى. لذلك فمن غير المناسب حذف الفقرات التي تختلف في المنحنى المميز لها، والادعاء بأن الفقرات المتبقية غير متحيزة لأن مجموعة الفقرات الأصلية أحادية البعد لكل مجموعة، فإن الفقرات المتبقية في الاختبار ما زالت تقيس سمات مختلفة لكلا المجتمعين الفرعيين وبذلك تكون متحيزة.. وفي المجال العملي لا يتوافر لدينا فقرات أحادية البعد بالتأكيد، لذلك فإن كان عدد قليل من الفقرات متباين عبر المجموعات الفرعية، وكان لهذه الفقرات معاملات تمييز منخفضة فمن المنطقي حذف مثل هذه الفقرات والادعاء بأن الفقرات المتبقية هي فقرات غير متحيزة.

تقنيات مربع كاي:

طور الباحثان شيونمان (Scheueman,1979) وكاميلي (Camilli,1979) تقنيات يمكن اعتبارها تقريب لطرائق نظرية الاستجابة للفقرة. والميزة الوحيدة لهذه الإجراءات أنها أبسط من الإجراءات المعتمدة على نظرية الاستجابة للفقرة. وتحدد الفقرة من خلال هذه التقنيات على أنها غير متحيزة إذا كانت نسبة الأفراد الذين أجابوا إجابة صحيحة على الفقرة هي نفسها ضمن فترة الدرجات نفسها في كلا المجتمعين الفرعيين. ويمكن اعتبار هذه الطرائق تقريبات لطرائق نظرية السمة الكامنة، وذلك لأن درجة الاختبار الملاحظة هي تعويض للسمة الكامنة عند تحديد الفقرة المتحيزة، ودرجة الاختبار الملاحظة هذه يمكن اعتبارها مقياس خاطئ للسمة الكامنة.

ويقسم تدريج الدرجة الملاحظة في تقنيات مربع كاي إلى عدة فترات، إذ تتم المقارنة بين

المجتمعات الفرعية في ضوء نسب الإجابة الصحيحة على الفقرة. فإذا تباينت النسب بين المجتمعات الفرعية فهذا يعد دليل على تحيز الفقرة. ويبين جدول (16-3) بيانات مثال قسم فيه تدريج الدرجات إلى أربع فقرات، تشير فيه الرموز N_{1j} و N_{2j} إلى عدد المفحوصين المتقدمين للاختبار في كلا المجموعتين الفرعيتين الأولى والثانية على التوالي في الفترة j، لذا فإن (25) مفحوص من المجموعة الفرعية الأولى و (315) مفحوص من المجموعة الفرعية الثانية لديهم درجات في هذه الفترة. وتشير الرموز O_{1j} و O_{2j} إلى عدد المفحوصين في المجموعات الفرعية الأولى والثانية وتقع درجاتهم في الفترة (J) وأجابوا على الفقرة إجابة صحيحة. وتحسب الكمية p_{1j} باستخدام المعادلة:

$$\frac{O_{1j}}{N_{1j}} = p_{1j} \quad \text{(5-16)................}$$

و p_{1j} هي نسبة المفحوصين في المجموعة 1 والفترة j وأجابوا على الفقرة إجابة صحيحة. فمثلاً للمجموعة الفرعية (1) وللفقرة (3) فإن نسبة الإجابة الصحيحة على الفقرة $P_{13} = \frac{23}{48} = 0.479$ وتحسب الكمية p_{2j} بالطريقة نفسها للمجموعة الفرعية (2). وتحسب الكمية P.J باستخدام المعادلة:

$$\frac{O_{2J} + O_{1J}}{N_{2J} - N_{1J}} = P.J \quad \text{(6-16)................}$$

وهي نسبة المفحوصين في الفترة j جميعهم وأجابوا على الفقرة إجابة صحيحة. وللفقرة الرابعة هذه النسبة تساوي:

$$2.99 = \frac{33 + 14}{92 + 65} = P.4$$

ويحسب إحصائي كاميلي (camilli's statistic) باستخدام الصيغة:

$$^{2}\chi_{J} \Sigma = \frac{^{2}(P_{2j} - P_{1j})\, N_{2j}\ \ N_{1j}}{(P._J - 1)\, P._J\, (N_{2j} _ N_{1j})} \sum_{1=J}^{J} = \chi_c^2$$

ويمكن اختبار الدلالة الإحصائية لـ $^{2}\chi_{2}$ بمقارنتها بـ 100 (1- α) وتوزيع مربع كاي وبدرجات حرية = J. وتمثل j هنا عدد الفقرات الكلي.

وبصورة متبادلة فإن مقدار $^2\chi_2$ يمكن اعتباره مؤشراً لمدى تحيز الفقرة. ويبين الجدول (16–4) حساب $^2\chi_2$ باستخدام بيانات الجدول (16-3). ويوجد نسخة من إحصائي كاميلي تحمل إشارة ويرمز لها بـ $^2\chi_2$ (إشارة)، ومع هذا الإحصائي فإن قيم $^2\chi_g$ تعطى إشارة موجبة أو سالبة على وفق قيمة p_{1j} إذا كانت أكبر أو أصغر من p_{2j}.

جدول (16-3): بيانات توضيحية لحساب $^2\chi_2$ و $^2\chi_s2$

الفقرة	مستوى الدرجة	N_1J	O_1J	P_1J	N_2J	O_2J	P_2J	$P.J$
1	13-14	25	22	0.880	315	300	0.952	0.947
2	12	24	18	0.750	110	99	0.900	0.873
3	10-11	48	23	0.479	118	93	0.788	0.698
4	9-1	65	14	0.215	92	33	0.358	0.299

مقتبس عن: Scheuneman - Anew method for assessing bias in test items. Journal of educationl measurement. 16, 143 - 152. copy right 1979 by the National council on measurement in education, washington, D.C. Adapted by permission.

جدول (16-2) توضيح حساب إحصائي كاميلي $^2\chi_2$:

الفقرة	التعويض لـ $^2\chi_j$	نتيجة الحساب
1	$\frac{^2(0.952 - 0.880)\ (315)25}{(0.947 -1)\ 0.947\ (315 + 25)}$	2.392
2	$\frac{^2(0.900 - 750)\ (110)24}{(0.873 -1)\ 0.873\ (110 + 24)}$	3.998
3	$\frac{^2(0.788 - 0.479)\ (118)48}{(0.698 -1)\ 0.698\ (118 + 48)}$	15.455
4	$\frac{^2(358 - 0.215)\ (92)65}{(0.299 -1)\ 0.299\ (92 + 65)}$	3.716

$$26.56\ = 3.716 + 16.455 + 3.998 + 2.392 = {}^2\chi_2$$

ويحسب إحصائي شيوني مان باستخدام الصيغة:

$$^2\chi_s = \sum_{1=J}^{J} \frac{^2(P._J - N_{1J} - O_{1J})}{P._J\ N_{1J}} + \sum_{1=J}^{J} \frac{^2(P._J N_{2j} - a_{2J})}{P._J\ N_{2J}} \quad \text{.................(8-16)}$$

ويبين جدول (5-16) خطوات حساب $^2\chi_s$، واقترح شيونمان أن توزيع $^2\chi_s$ يشـابـه توزيع $^2\chi$ بدرجات حرية عددها (J-1). وأشار العديد من الباحثين بأن هذا ليس صحيحاً إذ يمكن اعتبار $^2\chi_s$ مؤشراً لمقدار (لمدى) تحيز الفقرة، وهناك صيغة لـ $^2\chi_s$ تحمل إشارة موجبة أو سالبة تحسب من خلال إعطاء قيم $(P.J\ N.g- O_{1j})$و $(P._JN_{2J} -O_{2J})$ بناءاً على كون O_1J أكبر أو أقل من $P_{1J}N_{1J}$.

جدول (5-16): توضيح حساب إحصائي شيونمان $^2\chi_s$

الفقرة	التعويض للمجموعة الأولى	ناتج الحساب	التعويض للمجموعة الثانية	نتيجة الحساب
1	$\frac{^2[(25)(0.947)- 22]}{(25)\ (0.947)}$	0.118	$\frac{^2[(315)(0.947)- 300]}{(315)\ (0.947)}$	0.009
2	$\frac{^2[(24)(0.873)- 18]}{(24)\ (0.873)}$	0.415	$\frac{^2[(110)(0.873)- 99]}{(110)\ (0.873)}$	0.091
3	$\frac{^2[(48)(0.698)- 23]}{(48)\ (0.698)}$	3.293	$\frac{^2[(118)(0.698)- 93]}{(118)\ (0.698)}$	1.373
4	$\frac{^2[(65)(0.299)- 14]}{(65)\ (0.299)}$	1.519	$\frac{^2[(192)(0.299)- 33]}{(92)\ (0.299)}$	1.096

$$^2\chi_s = 1.096 + 1.373 + 0.009 + 1.519 + 3.293 + 0.415 + 0.118$$

$$^2\chi_s = 7.89$$

والقضية الحاسمة في استخدام أساليب مربع كاي تكمن في اختيار درجات القطع في تكوين الفترات، ويعد هذا قراراً مهماً إذ أن اختيار درجات القطع هذه يؤثر على قيمة كل من $^2\chi_c$ و $^2\chi_s$ (Ironson,1982) ويتم اختيار درجات القطع بحيث يكون عدد كافي من المفحوصين في كل فترة، وسيتم حالاً مناقشة النتائج التجريبية المتعلقة بعدد المفحوصين في كل فترة. وأشارت شيونمان (1979) إلى أنه ما لم تحدث إجابات خاطئة في كل فترة، فإن هذه الفقرة لا يمكن أن تساهم في اكتشاف التحيز، ومع ذلك فلم تخصص أي عدد الإجابات الخاطئة في كل فترة، وأشارت إلى أن طريقتها تعتمد بشكل خاص على اختيار فترات خاصة شريطة أن يتراوح عدد الإجابات الصحيحة الكلي لكل مجموعة مفحوصين ما بين 10 إلى 20 وذلك في كل فترة. ويشير ايرونسون إلى أن هذا المحك يجب أن يحقق تقنية كاميلي أيضاً. إضافة إلى ذلك فقد افترضت شيونمان أنه لاستخدام $^2\chi_s$ فإن الإجابات الصحيحة المتوقعة للمجموعة (ι) وفي الفترة (J)$(N_{iJ}\ \ P_{\cdot j})$ يجب ان يكون 5 على الأقل لكل i و J. وبطريقة مشابهة اقترح ايرونسون أنه لاستخدام $^2\chi_c$ يجب أن تكون التكرارات المتوقعة لكل من $N_{ij}\ P_{\cdot j}$ و N_{ij} (1-p) يجب أن تكون 5 على الأقل.

وإحدى مشكلات تقنية مربع كاي هو أن الدليل على تحيز الفقرة قد يكون مظهر زائف لخطأ القياس. وكما أشرنا سابقاً، ففي تقنيات مربع كاي يتم مقارنة المجتمعات الفرعية في ضوء نسبة الاستجابات الصحيحة على كل فقرة، وضمن كل فترة درجة. وإذا لم تكن المقارنة لكل فترة درجة فإنه يجب مقارنة نسب الإجابات الصحيحة على أنها قياس لصعوبة الفقرة، كذلك فإن الفروق في صعوبة الفقرة بين المجتمعات الفرعية ليس بالضرورة مؤشراً لتحيز الفقرة، فقد يعكس الفروق الحقيقية بين المجتمعات الفرعية على البناء الذي يقيسه الاختبار. ويعد إجراء مقارنات من هذا النوع محاولة لضبط مثل هذه الفروق. والمشكلة تكمن في أن مثل هذه الاداة تضبط في احسن الأحوال الفروق في الدرجة الملاحظة بين المجتمعات الفرعية. ومع ذلك فمن المرغوب به ضبط فروق الدرجة الحقيقية (أو مكافئها: الفروق في درجة السمة الكامنة) أكثر من فروق الدرجة الملاحظة. وعندما يتم ضبط الدرجات الملاحظة فقد توجد فروق في الدرجات الحقيقية بين المجموعات حتى للمفحوصين الذي يقعون في فترة الدرجة نفسها. وفي هذه الحالة فإن المجتمع الفرعي ذو الدرجات الحقيقة الأعلى سيجيب إجابة صحيحة على الفقرة. وهذا صحيح حتى للمفحوصين الذين درجاتهم الحقيقية متشابهة أو درجة السمة الكامنة لهم نفسها، فإن النسبة التي ستجيب بصورة صحيحة على الفقرة ستكون نفسها لكل مجموعة. وهنا قد تظهر تقنية مربع كاي مؤشر لتحيز الفقرة الذي هو صورة زائفة (مصطنع) لخطأ القياس.

مقارنة المقاييس التي تعتمد مربع كاي والسمة الكامنة لتحيز الفقرة

تحقق العديد من جماعة القياس من الارتباط بين مقاييس مربع كاي ونظرية الاستجابة للفقرة لتحيز الفقرة. وقد أجرى ايرونسون وسبكوفياك (Ironson & Sabkoviak, 1979) تحليلاً لخمسة اختبارات فرعية استخدمت في دراسة دولية طولية، وكان الارتباط بين إحصائي شيونمان $^{2}\chi_{s}$ و A_j = 0.485 (للبطارية الكلية) و 0.361 (للألفاظ) و 0.652 (للصورة – العدد) و 0.742 (لمجموعات الحروف) و 0.505 و (للحساب) و-0.047 (للمقارنات الفسيفسائية). ودون هؤلاء أيضاً الارتباطات بين $^{2}\chi_{s}$ (بإشارة) و A_g (بإشارة) وكانت للبطارية الكلية (0.575) وللمفردات (0.56)، وللصور – العدد (0.82) ولمجموعات الحروف (0.820) وللحساب (0.543) وللمقارنات الفسيفسائية (0.16) وقد لاحظ ايرونسون وسبكوفياك أنه من الصعب تفسير هذه الارتباطات لأنه غير معرف بمدى تطابق البيانات للنموذج ثلاثي المعلم في نظرية السمة الكامنة لكل مجموعة فرعية.وقد فحص كل من شيبرد وكاميلي واڤريل (Shepard, Camilli & Averill, 1981) الارتباط بين مقاييس مربع كاي والسمة الكامنة لتحيز الفقرة. وحلل هؤلاء البيانات لعينات من السود والشيكانو والبيض على اختبار لورج – ثورندايك للذكاء. (اللفظي وغير اللفظي) (المستوى 3 الصيغة B، 1954).

ويبين جدول(16-6) الارتباطات بين مقياس مربع كاي والسمة الكامنة لتحيز الفقرة.

جدول (16-6): الارتباطات بين معاملات السمة الكامنة ومربع كاي لتحيز الفقرة.

المجموعات المقارنة	مقياس مربع كاي	الاختبار الفرعي ونوع المعامل			
		اللفظي		غير اللفظي	
		بإشارة	بدون إشارة	بإشارة	بدون إشارة
سود – بيض	شيونمان	1.59	0.45	0.63	0.44
	كاميلي	0.68	ب40.	0.66	0.41
شيكانو – بيض	شيونمان	0.58	0.40	0.76	0.28
	كاميلي	0.68	0.37	0.76	0.03

أ. الارتباط بين $^{2}\chi_{s}$ (بإشارة) و A_g (بإشارة).
ب. الارتباط بين $^{2}\chi_{c}$ و A_g.

مقتبس عن: L. Shepard, G. Camilli and MiAverill, comparison of procedures for detecting test - item bias with both internal and external ability criterion: Journal of Educational Statistics, 6,317 - 376. copy right 1980 by the American Educational Research Association, Washington, D.c. Adapted by permission.

وقد أجرى الباحثان جيتسون ونايت (Getson & knight, 1980) دراسة محاكاة لمجموعتين من المفحوصين عددهم (1200) مفحوصاً أجابوا على (5600) فقرة، وحلل هؤلاء البيانات بالنموذج ثلاثي المعلم في نظرية السمة الكامنة.

وأجريت المحاكاة بحيث تقيس الفقرات السمات الكامنة المختلفة لكلا المجتمعين.ويستخدم هذا الموقف بالتحديد تحيز الفقرة. وبما أن البيانات من نوع المحاكاة فقد تمكن روندر وزملاؤه من حساب القيم الحقيقية لـ A_g وحسبوا ارتباطها مع القيم المحسوبة لكل من A_g و $^2\chi_s$.

وكانت قيمة معامل الارتباط الأول(0.80) والآخر (0.73). ويقترح بناءاً على هذه القيم أنه عندما تكون الفقرات أحادية البعد لكلا المجتمعين الفرعيين وفي حالة وجود تحيز الفقرة، فإن $^2\chi_s$ يكون دقيقاً كدليل على مقدار التحيز كما هو الحال لـ A_g. وقد دون هؤلاء معامل ارتباط (0.73) بين A_g و $^2\chi_s$.

إن تناول مجموعة معاملات الارتباط من الدراسات الثلاثة يصعب تفسيرها نوعاً. فقد كان تحيز الفقرة واضحاً فقط في دراسة روندر وزملاؤه. فقد كان الارتباط بين القيم المحسوبة لـ $^2\chi_s$ و A_g الحقيقية كبيراً كما هو الحال للارتباط بين القيم المحسوبة والحقيقية لـ A_g. وهذا يقتضي إمكانية استبدال$^2\chi_s$ بـ A_g.

وتقترح أن معاملات الارتباط في دراسة ايرونسون وسبكوفياك ودراسة شيبرد تؤشر قياس شيء مشترك لمقاييس مربع كاي ومقاييس المساحة، ولكن لا يمكن اعتبار مقاييس مربع كاي البديل المناسب لمقاييس المساحة. والمشكلة مع الاستنتاج الأخير تكمن في عدم وضوح وجود تحيز فقرة فعلياً في الاختبارات التي تمت دراستها من قبل ايرونسون وسبكوفياك وشيبرد. لذلك فإن ما تقيسه A_g و A_g (بإشارة) غير واضح. لذا فإن غياب قابلية تعويض مقاييس المساحة بمقاييس مربع كاي المبين في هذه الدراسات قد يكون مهماً أو غير مهماً.

التقنيات المعتمدة على صعوبة الفقرة:

لقد تم تطوير طرائق عدة لاستخدام صعوبة الفقرة في دراسات تحيز الفقرة. وتستخدم هذه الطرائق إما نسبة الإجابات الصحيحة (P) أو إجراء تحويل ما لها.

وهذه الطرائق هي تنوع للموضوع والذي يعتمد على واحد من تعريفين لمجموعة الفقرات غير المتحيزة.

والتعريف الأول هو أن مجموعة الفقرات تكون غير متحيزة إذا كانت صعوبات الفقرة للمجتمع الفرعي الثاني مرتبطة تماماً مع صعوبات الفقرة للمجتمع الفرعي الأول. فكر في

رسم خط بياني لصعوبات الفقرة للمجموعة الفرعية الثانية على المحور العمودي y وللمجموعة الفرعية الأولى على المحو الأفقي x، ويتطلب هذا التعريف أن تقع النقاط جميعها على خط مستقيم. ويجب ملاحظة أن هذا الشرط ينطبق على صعوبات الفقرة المحسوبة لكلا المجتمعين الفرعيين. ولصعوبات الفقرة المحسوبة من غياب كلا المجتمعين الفرعيين يمكن توقع مدى من التشتت حتى عندما تكون مجموعة الفقرات غير متحيزة وفقاً للتعريف الأول.

وقد تكون الطريقة المشهورة التي تعتمد هذا التعريف تلك التي طورها انغوف وفورد**(Angoff &Ford, 1973)** . وتستخدم هذه الطريقة مقياس الفرق لصعوبة الفقرة (delta measure) ويحدد الفرق للفقرة (g) على النحو الآتي:

$\hat{\Delta}_g = 4\hat{Z}_g + 13$, $\hat{Z}_g$ هنا هي الدرجة الزائية التي تقطع النسبة ($\hat{P}_g$) في التوزيع المعياري الطبيعي. وتمثل $\hat{p}_g$ مقياس النسبة الصحيحة لصعوبة الفقرة. ولتطبيق الطريقة بالرسم، تحسب أولاً: قيم $\hat{\Delta}_g$ لكل فقرة ولكل مجموعة فرعية، ثم الرسم البياني المذكور أعلاه في التعريف. ويبين جدول (16-7) بيانات افتراضية تصف إجابات مجتمعات فرعية اثنتان على عشرين فقرة.وتشير الإحصائيات $\hat{\Delta}_g$, $\hat{Z}_{g1}$,$\hat{P}_{g1}$ إلى المجتمع الفرعي الأول، $\hat{\Delta}_g$, $\hat{Z}_{g1}$,$\hat{P}_{g2}$ إلى المجتمع الفرعي الثاني. ويبين الشكل (16-2) تمثيل بياني لـ A_{g2} مقابل A_{g1}. ومعامل الارتباط لهذا الخط البياني (=0.96) والذي قد يكون أساسياً. مع ذلك يوجد نقطتين يجب حذفها من المجموعة وهذه يمكن اعتبارها نقاط لفقرات متحيزة.

إن الطريقة المذكورة أعلاه في اكتشاف تحيز الفقرة تقتضي مطابقة الفقرات على خط مستقيم ثم حساب قيمة (d_g)، أي المسافة بين الفقرة (g) والخط المستقيم، ويتم مطابقة الخط المستقيم على التمثيل البياني على أنه محور أساسي في المخطط، ويحسب ميل هذا الخط باستخدام المعادلة (16 –2 أ). وتشير σ_1^2 هنا إلى تباين الفرق للمجموعة (i) و P_{12} إلى الارتباط بين الفرق للعينيتين. وفي مثالنا كانت σ^2_1 =2.68 و σ^2_2= 17.926 ، و P_{12} = 0.961 ، وحسبت k وكانت تساوي (0.936) والقاطع m =$\bar{\Delta}g_2$ k- $\bar{\Delta}g_1$ حيث تمثل $\bar{\Delta}g_1$ - $\bar{\Delta}g_2$ متوسطات $\hat{\Delta}g_1$ و $\hat{\Delta}g_2$. وفي مثالنا كانت Δg_2 = 11.851 وΔg_1 =12.063 و m = (0.558) والقيمة المطلقة للمسافة أو بُعد النقطة التي تمثل الفقرة g من المحور الأساسي هي:

$$dg = \frac{m + \hat{\Delta}g_2 - \Delta g_1 \hat{k}}{\sqrt{1 + k^2}} \qquad (9\text{-}16)$$

والفقرات التي لها قيمة d_g كبيرة يكون انحرافها كافياً عن الخط لاعتبار الفقرة متحيزة،

ولسوء الحظ لا توجد قاعدة تجريبية (أو بالاستناد إلى نتائج التجربة) لتحديد متى يكون انحراف d_g كافياً ليؤشر تحيز الفقرة. ويبين الجدول (16-7) قيم d_g للفقرات العشرين. ومن فحص هذا الجدول يتبين أن d_{1g} = –2.170 و d_{20} = -2.31 وهذه القيم أكبر من قيم d_g للفقرات المتبقية، وهذا يقترح تحيز هذه الفقرات. وعندما تكون قيم عدة لـ d_g متقاربة (ليست أكبر بوضوح) ليصبح تفسير d_g أكثر صعوبة.

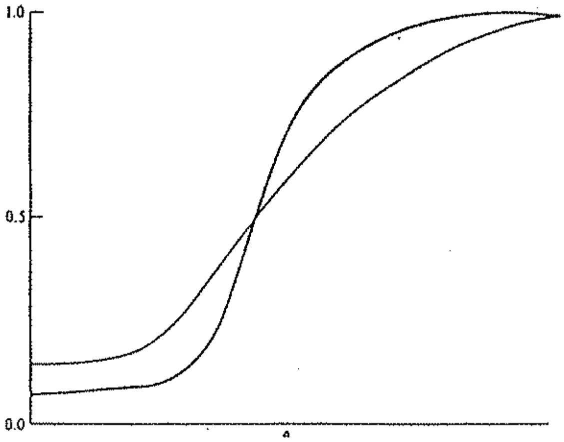
إحدى الصعوبات التي تواجه طريقة الفرق لصعوبة الفقرة هو أنه على الرغم من كون مجموعة فقرات غير متحيز. وفقاً لتعريف السمة الكامنة فإن طريقة مقياس الفرق قد تؤشر وجود تحيز. فعلى سبيل المثال، عند التفكير بالحالة التي تكون السمة الكامنة فيها موزعة اعتدالياً في كل مجموعة فرعية، وكان المنحنى المميز للفقرات على وفق نموذج المنحنى الطبيعي ثنائي المعلم، فقد اظهر هنتر(Hunter, 1975) أنه ما لم يكن معلم التمييز (a_g) متساوٍ للفقرات كلها فإن مخطط مقياس الفرق سيظهر بعض الدليل على تحيز الفقرة، بالإضافة إلى أن الفقرات ذات التمييز العالي ستبدو هي الأكثر تحيزاً. وتظهر هذه المشكلة عندما يكون توزيع السمة الكامنة ليس نفسه للمجتمعات الفرعية.وفي هذه الحالة يمكن حل المشكلة إذا تمت المطابقة بين المجموعات الفرعية في ضوء السمة الكامنة أو ما يكافئها بدلالة الدرجات الحقيقية، ومن الواضح أن هذا مستحيل. لذا فإن العديد من الباحثين اقترحوا أنماطاً من المزاوجة والتي هي تقريب للمطابقة على درجات السمة الكامنة.

جدول (16-7): مقارنة معاملات صعوبة الفقرات لمجموعتين:

dg	$\hat{\Delta}g_2$	$\hat{Z}g_2$	$\hat{P}g_2$	$\hat{\Delta}g_1$	$\hat{Z}g_1$	$\hat{P}g_1$	الفقرة
0.45-	12.466	0.133-	0.447	12.058	0.325-	0.407	1
0.29-	9.988	0.752-	0.226	9.648	0.837-	0.201	2
0.06-	15.813	0.703	0.758	16.209	0.802	0.788	3
0.33	11.620	0.344-	0.365	12.293	0.176-	0.430	4
0.01	18.578	1.394	0.918	19.265	1.566	0.941	5
0.18-	12.036	0.241-	0.404	11.993	0.251-	0.400	6
0.02	6.942	1.514-	0.065	6.845	1.538-	0.062	7
0.63	1.273	2.931-	0.001	1.688	2.828-	0.002	8
0.01	11.669	0.332-	0.369	11.888	0.277-	0.390	9
0.00-	0.730	2.567-	0.005	2.313	2.671-	0.003	10
0.22-	11.470	0.382-	0.351	11.331	0.417-	0.338	11
0.04	14.365	0.341	0.633	14.807	0.452	0.674	12
0.46	9.391	0.902-	0.183	10.111	0.722-	0.235	13

0.13	15.453	0.613	0.730	16.108	0.777	0.781	14
0.33	12.427	-0.143	0.443	13.166	0.041	0.516	15
0.04	14.692	0.423	0.663	15.152	0.538	0.704	16
0.22-	13.436	0.109	0.543	13.433	0.108	0.543	17
0.18-	12.860	0.034-	0.486	12.877	0.030-	0.488	18
2.71-	15.360	0.590	0.722	11.850	0.287-	0.387	19
2.31-	14.449	0.362	0.641	18.221	1.305	0.804	20

supgroup 2



supgroup 1

شكل (16-2) : مخطط بياني للفرق لمجموعتين فرعيتين

ويقترح انغوف (Angoff,1982) أن تكون المطابقة على المتغيرات الدخيلة على الاختبار، وترتبط بدرجاته. ويقترح جنسن (Jensen, 1980) استخدام الدرجة الكلية للاختبار، والاحتمال الآخر هو استخدام تقديرات الدرجة الحقيقية. وقدم عمل هنتر وكوهين (Huntr&Cohen, 1974) دعم نظري لهذا البديل. ويجب أن يبقى حاضراً نقطتين مهمتين عند استخدام مجموعات المطابقة أو المزاوجة.

الأولى: المطابقة بين المفحوصين على وفق المتغيرات الدخيلة والدرجات الملاحظة أو تقديرات الدرجات الحقيقية والتي تبقى مختلفة عن درجات السمة الكامنة. ونتيجة هذا فقد يظهر دليل مصطنع لتحيز الفقرة.

الثاني: أن محك عدم تحيز الفقرة يجب أن يكون للمجموعات المتطابقة عنه للمجموعات غير المتطابقة. وبما أنه للمفحوصين المتطابقين بدرجات السمة الكامنة تكون الفقرات غير المتحيزة

متساوية في صعوبتها. لذا فللمجموعات المتطابقة فإن ما يميز التمثيل البياني لمقياس الفرق أن لا يكون الاستنباط العالي مؤشراً على عدم التحيز. وبما أنه للمجموعات المتطابقة تكون الفقرات غير المتحيزة متساوية في صعوبتها فإنه يمكن فحص التحيز باختبار مكنمار لـ (Marscuilo &Mcsweeney, 1977) لتساوي نسب الأفراد للمتطلبات المناقشة أعلاه. والفقرات التي يكون إحصائي الاختبار الخاص بها دالاً إحصائياً تعد متحيزة. وكبديل يمكن أن نأخذ مقدار (كمية) الاختبار الإحصائي على أنه مقياس لمدى أو مقدار تحيز الفقرة.

طريقة أخرى لحل مشكلة اختلاف توزيعات درجة السمة الكامنة هو استخدام المجموعات المستعارة pseudogroups (Jensen,1980). وفي هذا التصميم تجري مزاوجة بين درجة كل مفحوص في المجموعة الفرعية الأصغر ودرجة مفحوص في المجموعة الفرعية الأكبر. وبعملية المزاوجة هذه تقسم المجموعة الفرعية الكبيرة إلى مجموعتين مستعارتين بحيث يكون توزيع درجات إحدى المجموعتين المستعارتين مطابق لتوزيع درجات المجموعة الفرعية الأصغر وتتضمن المجموعة المستعارة الأخرى بقية أفراد المجموعة الفرعية الأكبر. وبعدها نجري تحليل انغوف مرتين، مرة لكلا المجموعتين المستعارتين ومرة للمجموعتين الفرعيتين. وبعدها يمكننا مقارنة منحنى فرق الصعوبة لارتباطات المجموعات المستعارة مع المنحنى نفسه للمجموعات الفرعية. وتكون الفقرة متحيزة في حالة كون ارتباط المجموعات الفرعية أقل بشكل جوهري عن ارتباط المجموعات المستعارة.وثانية يوجد في هذا لتحليل تحذيرات عدة يجب أن تؤخذ بالحسبان.

الأول: أنه لا يوجد محك صعب وسريع لمقارنة ارتباطات المجموعتين. والثاني: لأن المطابقة لا تجري باستخدام الدرجات الحقيقية، فإن المطابقة بين المجموعتين المستعارتين من المتوقع أن تكون أكبر مما هو للمجتمعات الفرعية حتى عندما تكون الفقرات غير متحيزة بتعريف السمة الكامنة. ولسوء الحظ لا يتوافر مؤشر للفرق الذي يمكن توقعه.

ولاحظ منذ البداية أن طرائق صعوبة الفقرة في الكشف عن التحيز تعتمد على أحد التعريفين، وتعتمد طريقة انغوف على التعريف الأول. ويتطلب التعريف الثاني لمجموعة الفقرات غير المتحيزة تساوي فرق الصعوبة بين المجموعتين الفرعيتين على الفقرات جميعها. وبناءً على هذا التعريف فإن أي فقرة تظهر فرق صعوبة بين المجموعتين الفرعيتين أكبر في فرق الصعوبة المثالي لمجموعة الفقرات فإنها تكون متحيزة. ويجب ملاحظة أن أي فقرة تحقق محك عدم التحيز هذا في التعريف الثاني فإنها تحقق محك عدم التحيز في التعريف الأول، ولكن العكس ليس صحيحاً بالضرورة. ويستخدم أسلوب تحليل التباين المستخدم من قبل كل من كاردال وكوفمان (Cardall & Coffman, 1964) وكليري وهيلتون (Cleory &Hilton, 1968) وبليك وهوفر

(Plake & Hoover, 1977) هذا التعريف. للتعرف على هذه المنهجية انظر: دراسات هؤلاء. وأما أسلوب التمثيل البياني الذي يستخدم التعريف الثاني وصف من قبل اكترناكت (Echter-nacht,1974) وتخضع الطرق التي تعتمد التعريف الثاني للمشكلة نفسها كما هو الحال لتلك والتي تعتمد التعريف الأول. وخلاصة القول أن الأدلة على التحيز قد تكون زائفة بالنسبة لمستويات القدرة المختلفة للمجموعتين، والتصميمات لكل من مجموعات المزاوجة (المطابقة) والمجموعات المستعارة لها القدرة على الحد من هذه المشكلة، ولكن هذه القدرة محددة بالعوامل التي تمت مناقشتها.

مقارنة المؤشرات المعتمدة على معاملات السمة الكامنة ومنحنى الفرق

دوّن العديد من الباحثين الارتباطات بين d_g و A_g وكذلك بين الصيغ التي تحمل الإشارة، وكان الارتباط بين (الصيغ التي لا تحمل إشارة) عند ايرنسون وسبكوفياك (1979) d_g و A_g للبطارية الكلية (0.239) وللمفردات (0.373) و للصور - العدد (265.) ولمجموعات الحروف (-0.161) وللرياضيات (0.245) وللمقارنات الصورية (الفسيفسائية) (0.088)، بينما كانت الارتباطات المناظرة على الصيغة التي تحمل إشارة (491، 0.687، 0.675، 0.318، 0.419، 0.364). وبمقارنة معاملات الارتباط هذه مع المعاملات المناظرة لمربع كاي تبين بأن ثمانية معاملات أصغر ومعاملان اثنان في حدود التساوي ومعاملان اثنان أكبر. ودونت دراسة روندر وزملاؤه معامل ارتباط قيمته (0.61) بين القيم التقديرية لـ d_g والقيم الحقيقية لـ A_g ، ومعامل ارتباط قيمته (0.61) بين القيم التقديرية (dg) والقيم الحقيقية لـ Ag، ومعامل ارتباط قيمته (0.60) بين القيم التقديرية لـ d_g والقيم المحسوبة لـ A_g ، وهذه المعاملات أصغر من نظيراتها لمعامل شيونمان ($^2\chi_s$). أما في دراسة شيبرد وزملائه (1981) فقد كانت الارتباطات بين dg (بإشارة) و A_g (بإشارة) بين البيض والسود (0.52) على الاختبار اللفظي و (0.51) على الاختبار غير اللفظي، بينما الارتباطات المناظرة بين البيض والشيكانو فكانت (0.30) على الاختبار اللفظي و (0.29) على الاختبار غير اللفظي. وثانية فإن هذه أقل من الارتباط المناظر لطرائق مربع كاي، وباختصار نقول أن هذه تبدو اقل تبريراً لاستخدام d_g بدلاً من A_g أكثر مما هو لاستبدال $^2\chi_s$ بـ A_g.

مزايا الطرائق المختلفة ومساوئها:

إن الميزة الأساسية في استخدام نظرية السمة الكامنة أن مؤشرات تحيز الفقرة لا تظهر نتيجة الفروق في توزيع القدرة، وحتى عندما لا يكون هنالك تحيزاً في الفقرة فإن الفرق بين المنحنيات المميزة للفقرة تظهر عندما تقيس الفقرات أكثر من بُعد. وهذه تعد من المساوئ الواضحة للطرائق التي تعتمد نظرية السمة الكامنة، ويمكن التخلص من هذه السلبية من

خلال تطوير طرائق سمة كامنة متعددة الأبعاد. ويشترك في هذا العيب أيضاً طرائق مربع كاي وصعوبة الفقرة. وأما العيب الآخر فهو حجم العينة وعدد الفقرات اللازم لتطبيق طرائق السمة الكامنة. وكما أشرنا سابقاً فقد وجد هولين وزملاؤه (Hulin, etal, 1982) حد أدنى مقبول من الخطأ عن المعالجة بالنموذج ثلاثي المعلم لعينة حجمها (1000) مفحوص لـ (30) فقرة، ولعينة حجمها (500) مفحوص لـ (60) فقرة. وهذه المتطلبات يجب أن تتحقق في كل عينة لأن التقدير الدقيق مطلوب في كل عينة. أما النموذجين ثنائي المعلم وأحادي المعلم فإنها تتطلب عدد أقل من الفقرات والمفحوصين، فعلى سبيل المثال يوصي بيوريل (Burril, 1982) بعينة حجمها 500 مفحوص على الأقل لكل مجموعة، بينما يقترح رايت وستون(Wright &Stone,1979) عينة حجمها 200 مفحوص لكل مجموعة عند استخدام النموذج أحادي المعلم.

أما المزايا الأساسية التي تشترك بها كل من طريقة صعوبة الفقرة ومربع كاي فهي سهولة الحساب وأحجام عينة معتدلة نسبياً، إذ تعد العينة ذات حجم 100 أو 200 مفحوص في كل مجموعة أو عينة كافياً. أما العيب في طريقة صعوبة الفقرة فهي أن تحيز الفقرة يمكن أن يكون صورة مصطنعة (زائفة) لتباين التوزيع في المجتمعات الفرعية. وكمبدأ فإن هذه المشكلة يمكن الحد منها من خلال المزاوجة على درجات السمة الكامنة. وهذا المستحيل بعينه، إضافة على أن مزاوجة المفحوصين على متغيرات أخرى فشل كلياً في حل هذه المشكلة. ويمكن تصوّر تقنيات مربع كاي على أنها محاولات للمزاوجة على القدرة والمشاركة في مشكلة عدم المطابقة التامة مع طرائق صعوبة الفقرة.

الخلاصة:

في هذه الفصل تم تعريف الفقرات على أنها غير متحيزة إذا حققت المعيارين:

1/ تتأثر الدرجات على الفقرة في العينات الفرعية جميعها بمصادر التباين نفسها.

2/ وللمفحوصين المتجانسين على البناء الذي يقيسه الاختبار، تكون توزيعات مصادر تباين درجات الاختبار وغير المناسبة للبناء المقيس تكون نفسها في المجتمعات الفرعية جميعها.

وتم وصف طرائق ثلاثة للتحقق من تحيز الفقرة هي: نظرية الاستجابة للفقرة ومربع كاي وصعوبة الفقرة.

وتستخدم طرائق نظرية الاستجابة للفقرة مقارنة المنحنيات المميزة للفقرة للمجتمعات الفرعية وذلك باستخدام إحصائيات لمقارنة معالم الفقرة أو مقاييس المساحة بينما تستخدم طرائق مربع كاي المقارنات عبر العينات الفرعية لنسبة

الإجابة الصحيحة مع الدرجات الملاحظة في فترة درجات الاختبار نفسها. كذلك تم مناقشة طرائق مربع كاي المقترحة من قبل شيونمان وكاميلي. وتضمنت طرائق صعوبة الفقرة مقارنة مقاييس صعوبة الفقرة للعينات الفرعية ذات العلاقة أو تطابق المجموعات الفرعية، واحدة من كل مجموعة فرعية مناسبة.

ويمكن أن تؤشر الطرائق جميعها إلى تحيز الفقرة حتى عندما لا يكون موجوداً وفقاً للتعريف المذكور. ففي طرائق السمة الكامنة قد يؤدي تعدد الأبعاد إلى ظهور تحيز الفقرة (دليل خاطئ)، هذا بالإضافة إلى تداخل تعددية الأبعاد مع تحيز الفقرة. وعملياً نقول أن طرائق السمة الكامنة تعاني من الحاجة إلى بيانات لمجموعات كبيرة وبرامج محوسبة معقدة ومكلفة جداً. ويمكن استخدام طرائق مربع كاي وصعوبة الفقرة مع بيانات لمجموعات أصغر، ومنطقياً تكون أسهل في التطبيق. ومن هاتين الطريقتين يبدو أن منهجية مربع كاي واعدة بدرجة أكبر إذ أنها نظرية بدرجة أكبر بالإضافة إلى التعديلات التجريبية التي تجعل طرائق مربع كاي بديلاً لطرائق السمة الكامنة.

التمارين:

1/ فيما يأتي صعوبات 10 فقرات لمجموعتين من المفحوصين. باستخدام هذه المعاملات طبق طريقة انغوف وبيِّن الفقرات التي يظهر فيها تحيز.

صعوبات الفقرة لمجموعتين على الاختبار نفسه		
الفقرة	A	B
1	0.540	0.447
2	0.132	0.151
3	0.629	0.263
4	0.308	0.207
5	0.085	0.046
6	0.020	0.018
7	0.273	0.421
8	0.978	0.976
9	0.833	0.733
10	0.040	0.018

2/ لكل فترة من فترات الدرجات الأربعة للاختبار، يبين الجدول الآتي عدد المفحوصين (N_{2J}, N_{1J}) وعدد المفحوصين الذين أجابوا على الفقرة إجابة صحيحة. (O_{2J}, O_{1J}) لكلا المجموعتين. استخدم هذه البيانات لإجراء طريقة شيونمان وكاميلي. واستخدم القيمة الحرجة الملائمة لتحديد ما إذا وجد مؤشر الدلالة الإحصائية لتحيز الاختبار.

تكرارات عدد الذين اجابوا اجابة صحيحة والتكرار الكلي للمفحوصين لكل من المجموعتين والتي صنفت الى اربعة فقرات وفق درجة الاختبار.

وفق	أربعة فترات وفق درجة الاختبار			
الفقرة	O_1J	N_1J	O_2J	N_{2J}
1	16	56	36	100
2	31	47	48	72
3	41	53	44	62
4	40	43	33	35

3/ طبق اختبار مؤلف من 25 فقرة على عينة مؤلفة من 199 من الذكور و 269 من الإناث. وفيما يأتي التوزيع التكراري لكل مجموعة من الجنسين. كذلك دون عدد الذين أجابوا إجابة صحيحة على كل فقرة من الجنسين. استخدم هذه البيانات في إعداد فترة درجة مناسبة ومن ثم قم بتطبيق طريقة كاميلي. واستخدم القيمة الحرجة المناسبة لتحديد وجود تحيز للجنس دال احصائيا.

التكرار الكلي وتكرار الإجابات الصحيحة للمفحوصين مصنفة على وفق الدرجة الكلية والجنس.

الدرجة	الجنس			
	ذكور		اناث	
	التكرار الكلي	تكرار الإجابات الصحيحة	التكرار الكلي	تكرار الإجابات الصحيحة
5	1	1	0	0
7	0	0	1	1
8	2	1	2	1
9	2	0	2	2
10	4	2	7	4
11	4	2	5	2
12	4	4	7	2
13	4	3	7	6
14	5	2	18	9
15	11	7	28	14

15	23	7	19	16
15	35	15	25	17
19	37	12	22	18
18	34	12	27	19
19	38	14	26	20
12	19	15	24	21
13	15	8	13	22
1	1	6	6	23

الدرجات غير المذكورة لها تكرار = صفر لكلا الجنسين.

4/ تم احتساب معامل الصعوبة لراش والخطأ المعياري لصعوبة الفقرة لعينتين منفصلتين من الذكور والإناث.وتم تدريج معالم الفقرة لكلا المجموعتين بتحديد متوسط صعوبة الفقرات إلى صفر وقيم الصعوبات والأخطاء المعيارية مدونة في الجدول الآتي. استخدم هذه النتائج لتحديد أي الفقرات متحيزة (إن وجدت). (قد تفترض لتحقيق غرض هذا السؤال، أنه يوجد مؤشر كافٍ لاستنتاج أن الفقرات أحادية البعد لكلا الجنسين).

معاملات صعوبة راش (b) والأخطاء المعيارية للذكور/الاناث				
الاناث		ذكور		
SE	b	SE	b	الفقرة
0.138	0.007-	0.169	0.036-	1
0.252	2.071-	0.369	2.531-	2
0.170	2.564	0.679	2.367	3
0.130	0.511	0.154	0.758	4
0.341	2.831-	0.335	2.284-	5
0.158	0.653-	0.235	1.290-	6
0.130	0.528	0.160	0.341	7
0.202	1.466-	0.235	1.290-	8
0.165	2.452	0.29	2.832	9
0.131	1.218	0.156	1.469	10
0.129	0.846	0.154	0.829	11
0.155	0.579-	0.188	0.510-	12
0.246	2.009-	0.310	2.077-	13
0.133	1.357	0.154	1.102	14
0.135	0.142	0.162	0.238	15

الوحدة الخامسة

درجات الاختبار وتفسيرها

الفصل السابع عشر

17

التصحيح للتخمين وطرائق التصحيح الاخرى

الفصل السابع عشر

التصحيح للتخمين وطرائق التصحيح الاخرى

في طريقة التصحيح المتعارف عليها للاختبارات الموضوعية تكون علامة المفحوص عبارة عن مجموع علاماته على الفقرات التي تقدم إليها، إذ يُعطى المفحوص علامة واحدة لكل إجابة صحيحة، وصفر لأي إجابة أخرى. ويطلق على هذا أحياناً "عدد الإجابات الصحيحة". ويمكن التعبير عن قانون التصحيح هذا على النحو:

$$X_a = \sum_{i=1}^{n} \xi_{ai} \qquad (1\text{-}17)$$

حيث x_a: العلامة الكلية للمفحوص a.

x_{ai}: علامة المفحوص a على الفقرة I

وعندما تكون الفقرات ثنائية، لتصحيح فإن قيم ξai تكون (صفر أو واحد). ولا تعطى أية علامات تشجيعية عندما يحذف المفحوص الفقرة أو يترك مكان الإجابة فارغاً، وتكون أوزان الفقرات جميعها متساوٍ. ومع أن قانون التصحيح هذا بسيط ومباشر إلا أنه يواجه مشكلات عند استخدامه مع فقرات من نوع الاختيار من متعدد أو الصح – الخطأ.

ومنذ زمن بعيد أدركنا التمايز والتنوع في رغبات المفحوصين تجاه حذف الفقرات التي لا يعرفون الإجابة الصحيحة عليها، وهذا يسهم في تباين الدرجات الملاحظة، ومثل هذه التباين غير مرتبط بتباين المفحوصين على السمة المقيسة. افترض كمثال أن فيليس وجورج تقدما لامتحان قبول دراسة متخصصة، وكان الاختبار من نوع الاختيار من متعدد (100 فقرة). ويعرف كل من فيليس وجورج إجابات 60 فقرة من الاختبار، إلا أن فيليس ترك الفقرات المتبقية (وعددها 40) دون إجابة، في حين خمّن جورج إجابات الفقرات المتبقية. وفي حالة كون التخمين عشوائياً فقط فإن جورج سيجيب على 10 فقرات إجابة صحيحة وبالصدفة فقط إذا كان لكل فقرة أربعة بدائل إجابة، وفي حالة تطبيق طريقة التصحيح المتعارف عليها فإن جورج سيحصل على علامة كلية = 70 نقطة وفيليس60 نقطة فقط. وهنا يكون الفرق في الدرجات الملاحظة لا علاقة له بالفروق على السمة المقيسة، لكنه دالة للميل إلى التخمين

والصدفة العشوائي. وهذا يبرز أهمية استخدام بدائل طريقة التصحيح المتعارف عليها من قبل الفاحص الذي سيستخدم درجات اختبار واحد في اتخاذ قرارات مهمة مثل اختيار المتقدمين لوظيفة معينة، أو القبول في كلية أو عضوية التأهيل في تخصص معين.

صيغة التصحيح Formula Scoring

يكون الأسلوب التقليدي لتصحيح علامات الاختبار من أثر التخمين التفاضلي عبر المفحوصين باستخدام صيغة تصحيح. فقد لاحظ كل من رولي وتراوب (Rowley &Traub,1977) إن صيغة التصحيح تعتمد على نموذج يأخذ بعين الاعتبار مواقف محتملة ثلاثة:

- يعرف المفحوص بديل الإجابة الصحيح ويختاره.
- يحذف المفحوص الفقرة .
- يخمن ويختار عشوائياً أحد بدائل الإجابة التي عددها k.

وبالاعتماد على نموذج التخمين العشوائي، فإن صيغة أساسية لتصحيح العلامات الخام من أثر التخمين تكون على النحو الآتي:

$$\frac{O}{K} + R = X_c \quad \text{(2-17)}$$

حيث : X_c هي الدرجة المصححة من أثر التخمين.

R : هو عدد الإجابات الصحيحة عند المفحوص.

O : هو عدد الفقرات المحذوفة.

K : عدد البدائل لكل فقرة (وعدد البدائل نفسه لفقرات الاختبار جميعها). ويوضح جدول (17-1) حساب العلامات المصححة من أثر التخمين باستخدام هذه الصيغة لثلاثة مفحوصين تتنوع سلوكاتهم التخمينية. ويلاحظ أن التصحيح بهذه الصيغة يزيد علامة المفحوص الملاحظة، وذلك بمكافئته على الإجابات المحذوفة، بناءً على افتراض مفاده أن المفحوص لو حاول الإجابة على هذه الفقرات فإن احتمال اختياره للإجابة الصحيحة = $^1/_k$ وذلك لأنه يفترض عشوائية التخمين للفقرات المحذوفة جميعها.

جدول (17-1): حسابات توضيحية باستخدام صيغة التخمين لثلاثة مفحوصين تساوت علاماتهم الخام، ولكن باختلاف مدى التخمين على اختبار مؤلف من 20 فقرة لكل منها أربعة بدائل.

المفحوص	عدد الإجابات الصحيحة	عدد الفقرات المحذوفة	عدد الفقرات الخاطئة	Xc	X^1c
بين . ر	14	5	6	$14.0 = \frac{\text{صفر}}{4} + 14$	$12 = \frac{6}{3} - 14$
لويس. م	14	6	صفر	$15.5 = \frac{6}{4} + 14$	$14 = \frac{\text{صفر}}{3} - 14$
تامي. د	14	3	3	$14.75 = \frac{3}{4} + 14$	$13 = \frac{3}{3} - 14$

صيغة ثانية تعرف بـ تصحيح الصح مطروحاً منه الخطأ (Rights minus wrongs correction) وهي:

$$R - w/(k-1) = X_c^1 \qquad (3\text{-}17) \ldots\ldots\ldots\ldots\ldots\ldots$$

حيث X_c^1 : العلامة المصححة من التخمين.

w : عدد الإجابات الخاطئة.

k : عدد بدائل الفقرة.

ويبين جدول (17-1) أيضاً تطبيق هذه الصيغة على المفحوصين الثلاثة نفسهم. ومن الوهلة الأولى لا يظهر للعيان لماذا تعمل هذه الصيغة على أنها تصحيحية، والمنطق من مثل هذا الاستخدام هو حذف العلامات التي حسبت للمفحوص واكتسبها من التخمين العشوائي، إذ أن الافتراض الأساسي أن كل استجابة خاطئة تكون ناجمة عن التخمين العشوائي. لتوضيح كيف تعمل هذه الصيغة افترض أن لدينا اختبار مؤلف من فقرات لكل منها أربعة بدائل، أي أن k - 1 =3 وبنموذج التخمين العشوائي فإن w تمثل $^3/_4$ الفقرات التي لا يعرفها المفحوص فقط، وبقسمة w علىk-1 أو 3 ينتج تقدير لعدد الفقرات التي يحتمل أن تكون إجابة المفحوص عنها صحيحة بالتخمين، وتطرح هذه الكمية الأخيرة من درجة المفحوص عند إجراء التصحيح من أثر التخمين.

ومع أن المعادلتين 17-2 و 17-3 تؤديان إلى قيم مختلفة عددياً إلا أن ترتيب درجات

المفحوصين هو نفسه في كلا الحالتين. وفي الحقيقة يمكن تبيان أن المعادلة 3-17 عبارة عن تحويل خطي للمعادلة 2-17، لذلك فإنه لو طبقت كلا الصيغتين على المجموعة نفسها من الإجابات على الفقرات فإن الارتباط بينها يكون تاماً. ومع ذلك اقترح كل من تراوب وهامبلتون (Traub& Hambeltonm, 1972) أن التعليمات المرافقة لكلا الصيغتين قد تظهر عوامل سيكولوجية مختلفة لسلوكات المفحوصين على الرغم من التكافؤ الرياضي لهاتين الصيغتين.

الأساس المنطقي لصيغة التصحيح Rationale for Formula Scoring

يتفق مقترحو صيغة التصحيح على أن هذه الطريقة تزيد من ثبات الدرجات وصدقها، لأن الدرجة المصححة تعد تقديراً أفضل لدرجة المفحوص على السمة المقيسة من الدرجة الملاحظة غير المصححة. وقد تم تفسير المنطق النظري لهذا الجدال من قبل عدد كبير من السيكومتريين إلا أن لورد (lord,1975) قدم تفسيراً واحداً محكماً في محاولته لتوضيح النموذج الذي تتضمنه صيغ التصحيح، حدد فيه لورد قيمتين اثنتين لدرجة المفحوص هما x و Xc:

1/ x هي درجة المفحوص عندما أُعطي تعليمات ليجيب على الفقرات جميعها، واستخدم عدد الإجابات الصحيحة في إعطاء الدرجة.

2/ x_c هي درجة المفحوص عندما أُعطي تعليمات ليجيب على كل فقرة لديه معرفة جزئية كافية تمكنه من حذف بديل أو أكثر من بدائل الإجابة، وحذف الفقرات الأخرى، وحسبت الدرجة باستخدام المعادلة 2-17.

ويستند استخدام صيغة التصحيح على افتراض حاسم هو أن الفرق بين x و x_c للمفحوص نفسه وعلى الفقرات نفسها يُعزى إلى التخمين العشوائي فقط، والذي يؤثر على x بسبب التخمينات المحظوظة وغير المحظوظة عند إجبار المفحوص الإجابة على الفقرات التي تم حذفها حسب تعليمات صيغة التصحيح. وأشار لورد أيضاً إلى تساوي قيم كل من x و x_c لأي قيمة O :(O عدد الفقرات التي يجب حذفها ضمن شروط صيغة التصحيح). ومع ذلك فقد تختلف قيم x و x_c لمفحوص معين وفي اختبار ما، وذلك لأن تخمينات المفحوص الذكية قد تتجاوز $^{O}/_{k}$ أو قد تكون دونها ($^{O}/_{k}$ قيمة تعديل الدرجة الملاحظة باستخدام صيغة التصحيح). ولأن قيم x تتأثر بالتخمين العشوائي فإن تباين x المعاين يجب أن يكون أكبر من تباين المعاينة لـ x_c، ولأن تقديرات كل من x و x_c تعد تقديرات غير متحيزة للمعلم نفسه (درجة المفحوص الحقيقية)، فإنه يفضل التقدير ذو تباين المعاينة الأقل، لذا ففي النظرية يعد x_c تقديراً أفضل لقدرة المفحوص من x. ومع ذلك فإن الأبحاث التجريبية لم تؤدي إلى دعم هذا الاستنتاج.

الدراسات التجريبية لصيغة التصحيح:

في مراجعة للدراسات التي تقارن نتائج التصحيح المتعارف عليه، وجد دياموند وايفانز (Diamond&Evans, 1973) تشابه في تقديرات الثبات. أو أعلى قليلاً للدرجات غير المصححة، إلا أن معاملات الصدق كانت أعلى قليلاً للدرجات المصححة. وحتى لورد (Lord, 1963) استنتج أن التصحيح من أثر التخمين يؤثر في الصدق ويمكن تمييزه فقط عندما يكون هنالك تبايناً ملحوظاً في سلوكات المفحوصين التخمينية وبدائل إجابة عددها أقل من خمسة، واختبار صعب. لذا اقترح لورد (Lord, 1975) أن المفحوصين في الدراسات السابقة لم تقدم إليهم تعليمات حول متى يستخدمون التخمين ومتى يحذفون الفقرة في حالة التطابق مع افتراض التخمين العشوائي لنموذج صيغة التصحيح، وقد نادى بإجراء اختبارات تجريبية لافتراضات صيغة التصحيح وذلك بتطبيق الاختبار مع تعليمات صيغة تصجيح مناسبة، ومن ثم وعلى ورقة إجابة ثانية يطلب من المفحوصين الإجابة عن الفقرات التي حذفوها أولاً. فإذا كانت افتراضات صيغة التصحيح صحيحة فإن الدرجات الناتجة ضمن تعليمات صيغة التصحيح يجب أن لا تختلف عن تلك الموجودة على ورقة الإجابة الثانية والمصححة درجاتها من أثر التخمين. وأشارت الدراسات التجريبية المعتمدة على اقتراحات لورد (Bliss, 1980 , Cross & Frary, 1977) أنه عندما يجيب المفحوصين على الفقرات جميعها يحصلون على درجة خام أعلى مما لو أجابوا بناءاً على تعليمات صيغة التصحيح وصححت الإجابات كما اقترح لورد.

وتقترح النتائج أن سلوكات المفحوصين لا تتطابق مع افتراضات صيغة التصحيح. أحد التفسيرات المكنة، هو أن نموذج صيغة التصحيح لا يأخذ بالحسبان المعرفة الجزئية (أو طريقة استخدام المفحوص لها خلال العملية الاختبارية)، لذا فإن النموذج ضعيف رياضياً في تفسيره لسلوكات المفحوصين الواقعية (Gulliksn,1950; Lord & Novick,1968). ومع أن التحسين الكبير في التقدير الدقيق للدرجة المصححة يجب أن يظهر عند المفحوصين ذوي القدرة المتدنية الذين يحذفون فقرات عديدة (lord,1975) إذ أن هؤلاء قد يكونوا هم الأقل فهماً وتطبيقاً صحيحاً لتعليمات صيغة التصحيح، إلا أن هنالك تفسيراً أخراً محتملاً هو أن تعليمات صيغة التصحيح تولد عاملاً آخر في الموقف الاختباري والذي لا يمكن عمل شيء تجاهه لنطاق المهام التي يقيسها الاختبار، ومثل هذه التأثيرات تسمى الأخذ بالمجازفة- Risk (taking أو المراهنة الإرادية (Willingness To gample)

وتستخدم دراسات الفروق الفردية في الأخذ بالمجازفة تضمين الاختبار فقرات لا علاقة لها بموضوع الاختبار ولا يوجد لها إجابة صحيحة ضمن فقرات الاختبار، وعدد الفقرات هذه

التي يحاول المفحوص الإجابة عنها ضمن شروط التصحيح تعد مقياساً لميل المفحوص للأخذ بالمجازفة (لوصف هذه المنهجية أنظر إلى سلاكتر (Slakter,1968,1969) . وتشير ملخصات أبحاث دياموند وايفانز(Diamond & evans,1973) إلى أن الطلبة الأقل ميلاً للأخذ بالمجازفة يكونون أكثر ألتزاماً بتعليمات صيغة التصحيح أكثر من أولئك الذين ميلهم أكبر للأخذ بالمجازفة في الاختبارات الموضوعية.

ومثل هذه النتائج تبرز أسئلة حول قابلية الدرجات الناتجة عن شروط صيغة التصحيح للتفسير، وتبعاً لذلك فإن الكسب المتحقق في الصدق والثبات والوقت والجهد اللازمين والعلاقات العامة السلبية التي قد تنتج عن استخدام العقوبة على التخمين والتي تستخدم صيغة التصحيح التي يصعب الموازنة بينها وتعديلها.

صيغة التصحيح ونظرية الاستجابة للفقرة:

Formula scoring and item response theory

في وصف أكثر حداثة للورد (Lord , 1980) لكيفية أخذ صيغة التصحيح بعين الاعتبار في تقدير الدرجات الحقيقية في الاختبارات المطورة على وفق نظرية الاستجابة للفقرة (1). ويجب أن يسترجع القارىء من الفصل الخامس عشر أن ((p_g (θ)) يمكن تفسيرها على أنها احتمالية الإجابة الصحيحة لمفحوص مستوى قدرته (θ) على الفقرة (g). وفي الفصل نفسه لاحظنا أيضاً أن درجة المفحوص الحقيقية يمكن تقديرها بجمع الاحتمالات لفقرات الاختبار جميعها. وأشار لورد إلى أن هذا التطبيق قد يحتاج إلى تعديل فيما لو كان هنالك تفاضلاً بحذف الفقرات من قبل المفحوصين. وقد اقترح خطوات تحديد درجة عدد الإجابات الصحيحة للمفحوص (a) بالخطوات الآتية:

1/ حدد الفقرات التي أجاب عنها المفحوص (a)

2/ لكل فقرة منها، احصل على قيمة p_g (θ)، احتمالية أن يجيب المفحوصين الذين قدرتهم المحسوبة (θ) على الفقرة إجابة صحيحة.

3/ اجمع الاحتمالات جميعها.

وهذه العملية يرمز إليها بالصيغة:

$$\sum^{a} p_g(\theta) = \zeta a \qquad (4\text{-}17)................$$

1/ يمكن حذف هذا الجزء دون فقدان الاستمرارية مع الأجزاء التالية من هذا الفصل حيث ζa : الدرجة الحقيقية لعدد الفقرات الصحيحة.

$\sum^{a}$: تعني الجمع للفقرات التي أجاب عنها المفحوص a وتقدير الدرجة الحقيقية يتم تصحيحها فيما بعد من أثر التخمين بالصيغة:

$$\sum p_g^a(\theta) = \frac{\sum^a Q_g(\theta)}{k-1}\eta_a \qquad (5\text{-}17)$$

حيث $[1-p_g(\theta)] = Q_g(\theta)$

K: عدد بدائل الفقرة.

ويعتمد استخدام صيغة الدرجة الحقيقية في نظرية السمة الكامنة على افتراضين أساسيين هما:

1/ تعزى استجابات المفحوصين على الفقرات إلى مستويات القدرة في السمة الكامنة.

2/ يفهم المفحوصون تعليمات صيغة التصحيح بوضوح ويطبقونها، أي أنهم يحذفون الفقرة إذا (وفقط إذا) لا يوجد لديهم أي شيء أفضل من التخمين العشوائي ($^1/_k$) في اختيار الإجابة الصحيحة.

ولتوضيح استخدام المعادلتين (17-4) و (17-5) والشروط التي تعطي نتائج منطقية، دعنا نطرح الحالات المبينة في جدول (17-2) التي قد تظهر في تقدير الدرجات الحقيقية في نظرية الاستجابة للفقرة.

الحالة الأولى: لدينا أربعة مفحوصين بمستوى القدرة نفسه للسمة الكامنة (θ) وأربعة فقرات متكافئة في صعوبتها لكل منها أربعة بدائل، وعند هذه القيمة المعنية من (θ_2) احتمالية الإجابة الصحيحة على الفقرة (0.25). تمثل الحالة الأولى في جدول (17-2) النتائج التي يمكن الحصول عليها للدرجة الحقيقية ξa فيما لو لم يحذف المفحوصين الفقرات بالمعدل نفسه. من الواضح أن قيم ξa تتغير وتكون قيمتها أكبر للمفحوص الذي حذف أقل عدد من الفقرات. وهذا غير مرغوب فيه لأننا نعرف أن المفحوصين جميعاً بمستوى القدرة نفسه، ومع ذلك فإن صيغة الدرجات الحقيقية لهؤلاء المفحوصين الأربعة متماثلة مع أنهم أظهروا أنماط حذف استجابات مختلفة كثيراً.

الحالة الثانية: لدينا هنا أيضاً أربعة مفحوصين عند مستوى القدرة الكامنة نفسه، والفقرات غير متساوية في صعوباتها. فللمفحوصين عند مستوى القدرة هذا تكون احتمالية الإجابة الصحيحة على الفقرات (1 , 2 , 3 , 4) هي (0.25 , 0.25 , 50 , 0.50) على التوالي.

وتبين الحالة الثانية في جدول (17-2) قيم ζ_a و η_a لمفحوصين أنماط استجاباتهم المحذوفة مختلفة. وثانية نرى أن قيم ζ_a متغيرة وبتفضيل المفحوصين الذين حذفوا عدد أقل من الفقرات، ولكن قيم η_a للمفحوصين الأربعة متساوية.

الحالة الثالثة: وهنا لدينا أربعة مفحوصين بمستوى القدرة نفسه، وأربعة فقرات مختلفة في صعوباتها، يتبين من جدول (17-2) وفي الحالة الثالثة أن قيم ζ_a أعلى لمن حذف عدد اقل من الفقرات، ولكن قيم η_a غير متساوية لجميع المفحوصين، ويلاحظ أن للمفحوصين الأول والرابع قيم η_a منفردة لماذا؟ يتبين أن الافتراضات الأساسية للنموذج انتهكت من قبل هذين المفحوصين، فالمفحوص الأول أجاب عن الفقرة الأولى مع أن احتمالية الإجابة الصحيحة عنها أقل من ($^1/_4$)، والمفحوص الرابع حذف الفقرة الثالثة مع إن احتمالية الإجابة الصحيحة عنها تجاوز ($^1/_4$). لذا فإن هذين المفحوصين لم يستجيبوا بطريقة تتطابق مع صيغة التصحيح، فضلاً عن أن المفحوصين من مستوى القدرة نفسه أظهروا درجات مختلفة من الرغبة في الإجابة، لذا فإن الفقرات لا تقيس سمة مفردة.

وفي الحقيقة لا يمكننا معرفة الدرجات الحقيقية للمفحوصين، ويجب الاعتماد على القيم المقدرة لكل من:($\hat{\eta}_a$, $\hat{\zeta}_a$, θ_a)، كذلك لا يمكننا معرفة متى تنتهك الافتراضات اللازمة لتقدير صيغة الدرجة الحقيقية، وكما في صيغة تصحيح الاختبارات التقليدية، ومستخدمية هذه الطريقة تتطلب تطبيقات تجريبية، ويجب البرهنة على الفوائد العملية المستمدة من تطبيقاتها.

مكافئة المعرفة الجزئية:Awarding Credit For Partial Knowledge

بما أن الغرض العام لاستخدام صيغة التصحيح هو تجنب إعطاء المفحوصين درجات أو نقاط لا يستحقونها، فقد ظهرت مشكلة تصحيح أخرى أثارت انتباه ملحوظ وهي مكافئة المعرفة الجزئية. الأساس المنطقي لهذه الطرائق هو أن المفحوصين الذين يحصلون على درجات فقرة متماثلة في اختبار من نوع الاختيار من متعدد مصحح بالطريقة المتعارف عليها، فقد يختلف هؤلاء في درجات المعرفة (المعرفة الجزئية) المتعلقة بالفقرة. ويمكن تجميع طرائق التصحيح المصممة لنقل المعلومات حول المعرفة الجزئية بثلاث فئات عامة هي:

- وزن الثقة Confidence Weighting
- اجب حتى الصح Answer- until- Correct
- الوزن التفاضلي للفقرة Option Weighting

جدول (17-2): الاحتمالات التوضيحية لإجابات الفقرة الصحيحة
والدرجات الحقيقية لعدد الإجابات الصحيحة وصيغة الدرجات الحقيقية.

الحالات	الفقرات				ζ_a	η_a
	1	2	3	4		
الحالة الأولى						
المفحوص الأول	0.25	0.25	0.25	0.25	1.00	-
الثاني	0.25	0.25	0.25	-	0.75	-
الثالث	0.25	0.25	-	-	0.50	-
الرابع	0.25	-	-	-	0.25	-
الحالة الثانية						
المفحوص الأول	0.25	0.25	0.50	0.50	1.50	0.67
الثاني	0.25		0.50	0.50	1.25	0.67
الثالث	-	0.25	0.50	0.50	1.25	0.67
الرابع	-	-	0.50	0.50	1.00	0.67
الحالة الثالثة						
المفحوص الأول	0.10	0.25	0.50	0.75	1.60	0.80
الثاني	-	0.25	0.50	0.75	1.50	1.00
الثالث	-	-	0.50	0.75	1.25	1.00
الرابع	-	-	-	0.75	0.75	0.67

وهدف النقاش العام هنا ليس تبنياً أو دعماً لأحدى الطرائق الموصوفة، ولكن التعريف المختصر للقارئ مع تزويده بمراجع توضيحية لهذا المجال، والإشارة إلى الصعوبات التي قد تظهر عند محاولة تطبيق أحد بدائل التصحيح غير تلك المتعارف عليها.

وزن الثقة: Confidence Weighting

تم بناء صيغة التصحيح والتعليمات بحيث يؤشر المفحوص على درجة ثقته (تأكده) من

صحة كل استجابة. ومظهر محدد لطرائق وزن الثقة جميعها أن اختيار مفحوصين أثنين للإجابة نفسها قد لا يؤدي إلى حصولهم على الدرجة نفسها على الفقرة وذلك لاختلاف درجة الثقة بإجاباتهم. وقد راجع اكترناكت (Echternacht, 1972) دراسات وزن الثقة واستنتج أن مخطط وزن الثقة مع أنه يحتكم إلى أساس منطقي، فإن الزيادة المرغوبة في معاملات الثبات والصدق لا تتحقق عموماً باستخدام تقنيات أكثر تعقيداً سواء في طريقة تطبيق الاختبار أو تصحيحه. والمظهر المخيب للأمل هو أن عدد من دراسات وزن الثقة أظهرت بالفعل انخفاض قليل في معامل الصدق عند استخدام هذه الطريقة. ومن المهم ملاحظة إن تقنية التصحيح التي تستنفد بعضاً من وقت الاختبار ينتج عنها بالفعل زيادة جوهرية في ثبات الاختبار أو صدقه، مقارنة باختبار أطول يعطى بوقت اختباري مكافيء للحالة الأولى، لذا فإن الاهتمام بأبحاث تتعلق بهذا التطبيق ابتدأ منذ نهاية العقد السابق.

طريقة أخرى مقترحة لمكافئة المعرفة الجزئية هي الإجابة حتى الصح. أي أن المفحوص يقرأ فقرات الاختبار ويختار أحد بدائل الإجابة، وتعطى له تغذية راجعة فورية حول صحة انتقاؤه للإجابة، فإذا أختار الإجابة الصحيحة يعطى تعليمات بالانتقال إلى الفقرة التالية، وإن كانت إجابته خاطئة يعطى تعليمات لاختيار بديل إجابة آخر للفقرة نفسها. وتستخدم طريقة الاختبار هذه ورقة إجابة ذات خيال مخفي أو باختبارات محوسبة إذ تسجل عدد محاولات الإجابة لكل فقرة والطريقة التقليدية لتصحيح اختبار الاجابة حتى الصح هو طرح عدد الاستجابات المعطاة من المفحوص من العدد الكلي للاستجابات المكنة (Gilman & Ferry,1972) وأشارت اختبارات التحقق من ثبات الدرجات وصدقها المستخلصة من هذه الطريقة (Henna,1974&1975) إلى الارتفاع القليل لتقديرات الاتساق الداخلي عما هو في التصحيح المتعارف عليه، ولكن وجدت نتائج مختلطة في محاولات تطوير الصدق المرتبط بمحك. وظهر الاهتمام حديثاً بطرائق الإجابة حتى الصح من قبل ويلكوكس (Wilcox,1981,1982) الذي طور طريقة تصحيح لمنهجية الإجابة حتى الصح تعتمد على نموذج الدرجة الحقيقية القوي (Strong true Score Model)، واقترحت بعض النماذج للمواقف التي تحدد فيها درجة نسبة الإجابة الصحيحة على أنها نسبة الفقرات في الملف التي يستطيع المفحوص الإجابة عنها إجابة صحيحة. إن استخدام مثل هذا النموذج يتطلب افتراضات قوية حول توزيع الدرجات الحقيقية أو التوزيع الشرطي للدرجات الملاحظة (Lord & Novick 1968,ch. 23)، وأحد معوقات استخدام تصحيح الإجابة حتى الصح هو أن أداء بعض المفحوصين في الاختبار قد يتأثر تفاضلياً بعوامل مثل قلق الاختبار أو القدرة على التعلم أثناء العملية الاختبارية. واقترح حنا (Hanna, 1975) الحاجة إلى التحقق من هذه القضية.

وتستخدم الفئة الثالثة الواسعة من الطرائق الاختيار التفاضلي الموزون (Differential Option Weighting) الذي يعتمد على افتراض أن بدائل الإجابة على الفقرة تتباين في درجة صحتها، وأن المفحوص الذي يختار البديل الأكثر صحة تكون معرفته أكبر من المفحوص الذي يختار إجابة أقل صحة. ولا تتطلب هذه الطرائق تعليمات خاصة أو سلوكات غير مألوفة في تطبيق الاختبار فضلاً عن إعطاء درجات موزونة مختلفة لبدائل الإجابة تعتمد على درجة صحتها. وأحد الطرائق الشائعة للحصول على الدرجات الموزونة للبدائل تستخدم تقدير الخبراء أو المحكمين. فمثلاً استخدم باتنياك وتراوب (patniak & Traub) تقديرات الخبراء في ترتيب بدائل فقرات الاختبار جميعها على أساس صحتها النسبية، ثم استخدموا طريقة ثيرستون في المقارنات الزوجية لتحديد وزن كل بديل. وقد أعطى كل من ديفيس وفيفر (Davis &Fifer, 1959)وداوني(Downey ,1979) تعليمات للخبراء في تقديراتهم للبدائل على تدريج من (1 إلى 7) تمثل فيه الدرجة (1) إجابة خاطئة كلية و(7) إجابة صحيحة كلية، ثم احتسب متوسط تقديرات الخبراء واعتبر وزن للبديل. وتسمى هذه طرائق الأوزان النسبية (Rational weighting methods). كذلك وصف كل من جتمان (Guttman, 1941 وجوليكسن (Gulliksen,1950b) طريقة تجريبية لاشتقاق أوزان البدائل، وبصورة نموذجية يستخدم برنامج معين، ولكن المبدأ الأساسي هو أن البدائل التي يختارها المفحوصين ذات الدرجات الكلية الأعلى في الاختبار (أو على أي متغير محكي آخر) تحصل على أوزان أعلى من تلك المختارة من قبل مفحوصين درجاتهم الكلية أدنى . ونظرياً يرتفع معامل ثبات الفا للدرجات الموزونة المختارة بهذه الطريقة، ولكن نتائج الدراسات التجريبية لأوزان البدائل مخيبة للآمال عموماً. ومع ذلك هنالك زيادات متوسطة يتم الحصول عليها أحياناً في تقديرات الثبات، والزيادات في تقديرات الصدق كانت صغيرة جداً أو غير موجودة، إذ أن تكلفة استخدام مخططات تصحيح أكثر تعقيداً لا يمكن تعويضها بسهولة. وباهتمام خاص كانت نتائج هندركسون (Henderickson,1971) وداوني (Downey,1979) التي بينت أن أوزان البدائل نتج عنها زيادات في معاملات الفا الذي صاحبها اثر جانبي مشكوك فيه لمعاملات الصدق المرتبط بمحك. لذا ولغاية اليوم فإن الفوائد المتوقعة لأوزان البدائل ليست حقيقية، ومع ذلك فمن الممكن للتطبيقات المستقبلية لنظرية الاستجابة للفقرة أن نحصل على درجات موزونة للإجابات الخاطئة كما هو الحال للإجابات الصحيحة التي أثبتت أنها مثمرة. هذا وقد تم وصف دراسة واعدة في هذا المجال من قبل ثيسن (Thissen,1976). وفي مثل هذه الدراسات فإن التحسينات في المعلومات عن درجة الاختبار أو الدقة في تقديرات معالم الفقرة والقدرة يجب أن تؤخذ بعين الاعتبار كمحك في تقويم كفاءة أنظمة التصحيح الجديدة.

الخلاصة:

في التصحيح المتعارف عليه للاختبارات الموضوعية تعرف درجة الاختبار على أنها مجموع درجات الفقرات، وللاختبارات ثنائية التصحيح تكون درجة الفقرة صفر وواحد، وعندما يكون الفاحص مهتماً بمدى التخمين المؤثر في الدرجات يمكنه استخدام صيغة تصحيح. وتوجد طريقتان لصيغ التصحيح التي تعتمد على نموذج التخمين العشوائي الذي يفترض أن المفحوص يخمن عشوائياً عبر الاختيارات وعددها k للفقرة الواحدة. وعلى الرغم من المنطق النظري لصيغة التصحيح فان الدراسات التجريبية لم تدعم وجود زيادة جوهرية في ثبات درجات الاختبار أو صدقها. ومن الأسباب المحتملة هو أن السلوك في الموقف الاختباري يتأثر بسلوكات أخرى مثل الأخذ بالمخاطرة عند استخدام صيغة التصحيح في الاختبارات التقليدية.

وتعتمد إحدى طرائق حساب الدرجة الحقيقية على نظرية الاستجابة للفقرة والتي وصفت في هذا الفصل. واقترح لورد(1980) أن الصيغة قد تكون مفيدة في حالة حذف فقرات بأعداد مختلفة من قبل المفحوصين ذوي المستويات المتشابهة. ويمكن تمييز هذه الفوائد فقط عندما يطبق المفحوصين تعليمات صيغة التصحيح بصورة صحيحة، وتحقيق أحادية البعد والاستقلال المحلي وفق نماذج نظرية السمة الكامنة جميعها.

واقترحت طرائق أخرى لتصحيح الاختبارات تعتمد على افتراض أن المفحوصين لديهم درجات معرفة جزئية متباينة على فقرات الاختبار. وتتضمن هذه الطرائق أوزان الثقة، وطريقة الإجابة-حتى-الصح، واستخدام أوزان بدائل مشتقة تجريبياً أو منطقياً، والتي تعتمد على افتراض أن الاختيارات المتنوعة لفقرة اختيار من متعدد قد تختلف في درجة صحتها، وأن المفحوصين الذين يختارون بدائل مختلفة يكون لديهم معرفة جزئية مختلفة للمحتوى. ولغاية الآن، فإن الطرائق المختلفة أدت إلى زيادة قليلة مخيبة للآمال لثبات درجات الاختبار وأحياناً انخفاض في صدق درجات الاختبار.

التمارين:

1/ تتضمن المصفوفة التالية درجات خمس مفحوصين على 10 فقرات من نوع اختيار من متعدد:

المفحوص	1	2	3	4	5	6	7	8	9	10
أ	0	0	1	0	-	1	0	0	1	1
ب	1	1	1	1	1	1	1	1	1	1
ج	0	-	1	-	1	1	1	0	1	1
د	0	1	0	1	1	1	0	0	1	1
هـ	-	-	1	1	1	1	0	-	1	1

أ- احسب الدرجة الكلية لكل مفحوص على أساس عدد الفقرات التي أجابوا عنها إجابة صحيحة.

ب- هل يعد المفحوصين ج، د، هـ متساويين في الكفاءة للمحتوى الذي يقيسه الاختبار.

ج- ما الدرجات التي يحصل عليها المفحوصين فيما لو أجري تصحيح للدرجات من أثر التخمين والذي يقلل من درجة المفحوص الذي يستخدم التخمين

د- اثبت أن ترتيب المفحوصين يكون نفسه فيما لو طبق التصحيح الذي يكافيء المعرفة الجزئية للاستجابات المحذوفة.

2/ على افتراض أن باحث أهتم بمقارنة تقديرات الثبات للدرجات الناتجة عن صيغة التصحيح باستخدام بيانات الفقرة فقط مثل المبينة في التمرين الأول. أقترح طريقة مناسبة لتقدير ثبات كلاً من الدرجات التقليدية ودرجات الصيغة دون الحصول على بيانات أخرى.

3/ أهتم أستاذ فيزياء بخفض أثر التخمين على درجات الاختبار النهائي، وقرر تطبيق صيغة التصحيح. وعند فحصه لأوراق الإجابة لاحظ أن الطلبة أجابوا على فقرات الاختبار جميعها، ماذا سيكون أثر التصحيح من التخمين في هذا الموقف.

4/ باستخدام طريقة جديدة في تطبيق الاختبارات، والتي تتطلب من المفحوصين التأشير إلى درجة الثقة لاستجاباتهم، وجد الباحث أن تقدير الثبات للدرجات الموزونة بثقة كانت 0.85 ، في حين تقدير ثبات الفقرات نفسها المطبقة بالطريقة التقليدية المتعارف

عليها كان 0.70، واستغرق الاختبار 90 دقيقة للمجموعة التجريبية في حين استغرق 50 دقيقة للمجموعة الضابطة. فماذا تستنتج عن فعالية طريقة التصحيح التجريبية؟

5/ افترض أن بيانات الاستجابات المبينة في التمرين الأول جمعت لاختبار طور وفق نظرية الاستجابة للفقرة وبالنموذج اللوغاريتمي أحادي المعلم. وكانت درجة السمة الكامنة المحسوبة للمفحوصين (ج، د، هـ) نفسها وذلك من خلال دالة المنحنى المميز للفقرة، وكانت القيم المحسوبة لـ $p_g(\theta_i)$ لكل فقرة على النحو الآتي:

رقم الفقرة	1	2	3	4	5	6	7	8	9	10
$p_g(\theta)$	0.20	0.30	0.60	0.50	0.80	0.90	0.20	0.10	0.95	1.00

أ- ما تقديرات الدرجة الحقيقية للمفحوصين ج، د، هـ؟

ب- ما تقدير عدد الدرجات الصحيحة الحقيقية للمفحوصين الثلاثة؟

ج- ما صيغة تقدير الدرجة الحقيقية لهؤلاء المفحوصين؟

الفصل الثامن عشر

18

إعداد المعايير

الفصل الثامن عشر

إعداد المعايير

يتطلب العديد من المواقف تحديد درجة القطع قبل تفسير الأداء على الاختبار، على سبيل المثال: تقسم بعض البرامج التعليمية إلى جزأين، إذ يتقدم المفحوص لاختبار بعد إنهاء الجزء الأول ليسمح له بعدها بالتقدم إلى الجزء الثاني فقط إذا كانت الدرجة على الاختبار تساوي أو تتجاوز درجة القطع المحددة. وفي حالات أخرى مثل التأهيل لبعض التخصصات التطبيعية والتسكين يطلب من المتقدمين إكمال المعرفة المتخصصة في الاختبار. ويتم الموافقة على التأهيل فقط إذا تساوت درجة المتقدم ودرجة القطع أو تجاوزتها. وتطبق المواد درجات القطع تعرف عادة على إنهاء إعداد المعايير.

وعلى المدى الواسع يمكن القول بأن هناك ثلاثة أساليب لإعداد المعايير. تتطلب الأولى فحص محتوى الاختبار من قبل خبير أو أكثر والذي يصدر حكماً بالاعتماد على الانطباع الكلي عن محتوى الاختبار. ويعتمد الأسلوب الثاني على الحكم على محتوى الفقرات المنفردة. ووجهة نظر تقليدية لهذه الأساليب في إعداد المعايير هو أنها لا تستخدم القضايا السيكولوجية التقنية (Gallagher, 1978 & Linn, 1978) والأسلوب الرئيسي الثالث لإعداد المعايير يعتمد على أداء المفحوصين، ويعد هذا التوجه سيكومتري نوعاً ما . ومع ذلك فإنه يتطلب عنصر مهم من الأحكام العامة، إذ أن جميع الأساليب التي تصنف تحت هذا الأسلوب تتطلب من معد المعايير اختيار مجموعة مفحوصين ثم اختبار أداءها. وهذه الأساليب البراغماتية الثلاث في إعداد المعايير سيتم وصفها في الأجزاء التالية من هذا الفصل.

والطريقة الأكثر شيوعاً لاستخدام درجة القطع الناتجة عن تطبيق أحد أساليب إعداد المعايير هو تطبيقها مباشرة على الدرجات الملاحظة. بمعنى أن معد المعايير لو أستنتج أن درجة القطع المناسبة كانت 0.69 على سبيل المثال، بعدها على المفحوص أن يجيب على 69% أو أكثر من فقرات الاختبار إجابة صحيحة لاجتياز الاختبار. وطريقة بديلة لاستخدام درجة القطع هي تطبيق درجات النطاق. وفي هذا الأسلوب تعالج فقرات الاختبار على أنها عينة من مجال أكبر من الفقرات، وكل مفحوص يعد حاصلاً على درجة النطاق (أو الحقيقة) والتي تحدد عادة على أنها نسبة الفقرات من النطاق التي يستطيع المفحوص الإجابة عنها إجابة صحيحة. وتستخدم درجة القطع التي يضعها معد المعايير لتجزئة تدريج درجة النطاق إلى منطقتين، وهذه تسمى درجة قطع تدريج النطاق. وفي اختبارات التحصيل تسمى هذه المناطق

حالات السيطرة، فالمفحوصين الذين درجاتهم تساوي درجة القطع أو أعلى منها يطلق عليهم أسم مسيطرين حقيقيين، وأما المفحوصين الآخرين فيطلق عليهم أسم غير مسيطرين حقيقيين. وحالما يتم وضع درجة قطع على متصل النطاق، فإن المشكلة تكمن في تحديد درجة القطع الملاحظة التي تسمح لمستخدم الاختبار استخلاص الاستنتاجات الأكثر ملائمة لحالات السيطرة الحقيقية للمفحوص. ويمكن أن يرافق تحديد درجة القطع الملاحظة الأخذ بعين الاعتبار العلاقة بين الدرجات الملاحظة ودرجات النطاق والتي يرافقها خطأ تصنيف المسيطر الحقيقي على أنه غير مسيطر وغير المسيطر الحقيق على أنه مسيطر.

ويطرح هذا الفصل أساليب تحديد درجة القطع الملاحظة والقضايا المتعلقة بها. تحت عنوان الاعتبارات التقنية في إعداد المعايير.

وقبل الرجوع إلى النطاق الرئيسة في هذا النقاش يجب أن نعرج على نقطتين ثانويتين. الأولى أننا حددنا النقاش في إعداد درجة قطع واحدة تجزئ متصل الدرجة إلى منطقتين اثنتين بهدف الحفاظ على بساطة العرض ومن ناحية أخرى يبدو أن استخدام درجة نجاح واحدة هي الحالة الأكثر شيوعاً في التطبيقات العملية. وبغض النظر عن حاجة مواقف لإعداد معايير بعينها والتي تتطلب استخدام عدة درجات قطع من أجل تقسيم متصل الدرجة إلى عدة مناطق. كمبدأ، فإن جميع الأساس المعروضة في هذا الفصل يمكن تعميمها لتناول مشكلات تتطلب درجات قطع عديدة. والثانية: هي استخدام درجات القطع في اختبارات القبول والتوظيف عندما تستخدم الاختبارات في التنبؤ عن حالات السيطرة المطلقة على محك خارجي. وهذه الموضوعات عرضت في الفصلين الحادي عشر والثاني عشر ولن نتعرض إليها مباشرة في هذا الفصل.

الأساليب لإعداد المعايير:

مع أن أدب القياس وصف أكثر من 30 منهجية لإعداد المعايير في السنوات الأخيرة (Behuniak, Archam bault & Gable, 1982). إلا انه يمكن تصنيفها ضمن واحدة من فئات ثلاثة رئيسة (ويمكن للقارئ المهتم بمخططات تصنيفية إضافية الرجوع إلى (Glass, 1978, Hambelton, 1980 , Gaeger, 1979, Meskauskas , 1976 , Millman, 1973). وتستخدم المنهجيات الرئيسة المعتمدة هنا:

1/ أحكام تعتمد على انطباع شامل عن الاختبار أو ملف الفقرات

2/ أحكام تعتمد على محتوى فقرات الاختبار منفردة.

3/ أحكام تعتمد على أداء المفحوصين في الاختبار.

الأحكام المعتمدة على الانطباع الشامل:

هنا يفحص فريق من الخبراء محتوى الاختبار، واعتماداً على الانطباع الكلي ومجال المحتوى تقترح النسبة التي يجب أن تكون الإجابة عليها صحيحة، وذلك لمن سيكون له أدنى مستوى من الكفاءة ليكون أداؤه ضمن المستوى المطلوب للأداء وكل محكم عادة يضع معيار معين، وفي المعيار النهائي يؤخذ متوسط المعايير جميعها وقد اقترحت شبرد (Shepard, 1976) استخدام مجموعة أحكام متعددة تمثل وحدات متعددة لمن يهتم بنتائج الاختبار. فعلى سبيل المثال في شهادة الكفاءة الجامعية الأولية يمكن ضم فئات من الطلبة والمعلمين وأولياء الأمور وممثلين عن المجتمع المحلي على مدى واسع .

وتعتمد هذه الأحكام في بعض الأحيان على السماح بخطأ، وأشار جلاس (Glass, 1978) إلى هذا الاستخدام على أنه "العد العكسي الذي يبدأ من 100%" ، وهو أن معد المعايير يفترض مستوى مرغوب أو مفضل للأداء على الاختبار يجب أن تكون الإجابة على 100% منها إجابة صحيحة ولكنها تسمح ببعض الأخطاء ، الناتجة عن خطأ القراءة أو عدم وضع إشارة أو عدم الانتباه أو أخطاء التصحيح وهكذا.

فقد تستخدم بعض المعرفة عن مجتمع المفحوصين من قبل المحكمين. فعلى سبيل المثال أورد جلاس في تقرير عن طفل في الصفوف الابتدائية يفيد أنه يمكن تدريب هذا الطفل فقط لغاية 70% من الدقة في الجمع لمنزلة واحدة. ومن غير الشائع وجود معلمين يقترحون نسبة ما بين 60-70% على أنها معيار معقول لاختبارات من هذه النوعية، وقد يكون اقتراحهم هذا بسبب خبرتهم مع مفحوصين عند هذا المستوى.

ومع أن الأحكام الشاملة هي واحدة من المنهجيات الأكثر استخداماً في استخداماً في إعداد المعايير، فإنه من الصعب الدفاع عنها عن وجهة نظر منطقية أو سيكومترية ونقد شائع لها هو أن مطور الاختبار لا يمكنه معرفة ما إذا كانت عينة خبراء مختلفين لم تؤسس المعيار عند نقاط مختلفة. واحد الحلول المنطقية لهذه المشكلة تظهر الحاجة لإجراء دراسات مكررة .

وعلى افتراض أن العدد الكلي في الحكام المتوافر والمناسب ثابت، وفي حالة التكرار مرتين فإن عدد الخبراء لكل دراسة سينخفض إلى النصف، وبالتالي فإن مجموعة المعايير تعد من قبل عدد أقل من الخبراء، وهذه قد تؤدي إلى انحرافات من عينة لأخرى. أكثر مما هو للمعايير التي تعد من قبل عدد أكبر من الخبراء . ومشكلة أخرى أنه لا يوجد تأكيد بأن الخبراء المختلفين تعتمد على انطباعاتهم على مظاهر الاختبار نفسها أو أن تتشابه إدراكاتهم لمجال المحتوى، وهذه قد تؤدي إلى انحرافات غير مرغوب بها للمعيار من عينة خبراء لأخرى. وعند استخدام هذه المنهجية يجب على مطور الاختبار أن يوفر على الأقل توثيقاً لعدد الخبراء ومؤهلاتهم وعملية اختبارهم والتعليمات التي زودوا بها وتوزيع استجاباتهم.

الأحكام التي تعتمد على محتوى الفقرة:

تعد هذه المنهجية من منهجيات إعداد المعايير المنهجية التي درست على المدى الأوسع انتشاراً، وهناك ثلاثة طرائق مشهورة في هذه الفئة سيؤخذ بها، وبالترتيب التاريخي لها. فقد اقترح نيدليسكي (Nedelskyo, 1954) التقنية الأولى، والتي صححت على وجه الخصوص لفقرات من نوع اختبار من متعدد. وقد اهتم نيدلسكي بتأسيس معايير لأقل كفاءة أو تأهيل للمفحوصين في المستوى الجامعي.

وكانت المعايير التي حددها باستخدام هذه على النحو الآتي:

1/ يعطي كل محكم تعليمات (وعادة يكون خبير مؤهل في مجال المحتوى) لوضع إشارة لكل فقرة يبين عدد الاستجابات التي يمكن للمفحوص الأقل كفاءة (وهي لأقل طالب) أن يحذفها.

2/ يسجل الحكام ولكل فقرة مقلوب عدد الاستجابات . على سبيل المثال لفقرة من خمس بدائل، فلو استخرج بديلين فإن تسجيل قيمة المقلوب تكون $^{1}/_{3}$.

3/ إيجاد M وهي مجموعة المقلوب عبر الفقرات جميعها، ويمكن عدّها الدرجة المحتملة للمفحوص الأقل تأهيلاً، كما حددت من قبل المقدرين لذلك الحكم المفرد.

4/ تؤخذ متوسطات قيم M عبر جميع المتعلمين (μ_M) . وقد اقترح ندلسكي أما الدرجات الناجحة في الاختبار يجب أن تساوي $\mu_M + \sigma_M$ حيث k قيمة اصطلاحية تختار بشكل مناسب، وتتراوح قيمتها بين 0.5 و .10 وقد تم نقد الافتراض المنطقي المستخدم في اختبار K وتعديل قيمة μ_K ، لذلك فإن بعض مستخدمي هذه التقنية يفضلونه البساطة ووضع أقل نقطة نجاح عند μ_M . (Meskauskas,1976). واقترحت المنهجية الثانية من قبل انغوف (Angoff, 1971) ووصفها بالأساس في ملاحظة تذيلية وذلك في توضيحه لكيفية إجراء تحويلات للتداريج المختلفة بدون بيانات معيارية. وتعطي هنا للمحكمين تعليمات بالأساس للتفكير في المجموعة الأقل قبولاً، وكذلك تقدير نسبة المجموعة الأقل قبولاً، وكذلك تقدير نسبة المجموعة الأقل قبولاً (ولكل فقرة) والتي تستطيع الإجابة على الفقرة إجابة صحيحة (ويمكن التفكير بهذا على أنه احتمالية الإجابة الصحيحة على الفقرة لمن لديه الحد الأدنى المقبول من الكفاءة) وثم تجمع هذه الاحتمالات عبر الفقرات جميعها للحصول على أدنى درجة نجاح وضعت من قبل المحكم المفرد . والقيمة المتفق عليها لتقديرات المحكمين جمعيهم تعد أقل درجة نجاح.

واقترح أيبل (Ebel, 1971) نظام مشابه لا نخوض ولكنه ميز أن كل محتوى الفقرة المناسب ومستوى صعوبتها قد تؤثر على المحكمين حول الأداء المتوقع للمفحوص الأقل كفاءة على الفقرة. وتستخدم هذه التقنية شبكة ذات بعدين في تصنيف الفقرات (رباعية الخلايا) .

أحد أبعاد الشبكة هو مستوى قبول الفقرة، أما البعد الثاني فهو للصعوبة وعادة ما يكون ذو ثلاثة مستويات انظر الجدول (18–1) . وفي البداية تصنف الفقرات في خلايا الشبكة، وبعدها يؤشر المحكم النسبة المئوية للفقرات في الخلية ويجب أن يجيب عليها المفحوص الأقل كفاءة إجابة صحيحة. ويمثل جدول (18–1) نسب مئوية افتراضية في كل خلية والتي يجب الحصول عليها من قبل كل محكم. وكذلك يبين عدد الفقرات لكل خلية في الاقواس أسفل النسبة المئوية. وتحسب أدنى درجة نجاح موصى بها من قبل محكم واحد بالمعادلة:

(M) P 3 = Xc .... (1-18)

جدول (18–1) : جدول توضيحي ونسب الفقرات المصنفة من قبل محكم واحد باستخدام منهجية ايبل في إعداد المعايير لاختبار مؤلف من 200 فقرة

مستوى القبول	مستوى الصعوبة		
	سهل	متوسط	صعب
اساسي	%90	%50	%10
	20 فقرة	(25)	(5)
مهم	%60	%30	%20
	(35)	(22)	10
مقبول	%40	%20	%10
	(19)	(12)	(15)
عليه استفسار	%25	-	-
	(7)	(20)	(10)

0.20 + (19) 40 + (10)0.50 + (22)0.30 + (35)0.6 +(5)10 + (25)0.50 + (20) 0.90+ =Xc
.73.85 =(10)0 + (20)0 + (7)0.25 + (15) 10 + (12)

حيث Xc درجة القطع بدلالة الدرجات الخام، و p نسبة الفقرات في الخلية التي يجيب عنها المفحوصين الأقل كفاءة.

وM هي عدد الفقرات في الخلية

والمجموع هو المجموع الكلي للخلايا الاثني عشر

ويمكن الحصول على درجة النجاح النهائية بحساب متوسط قيم Xc لجميع المحكمين. ويمكن تحويل هذه القيمة إلى نسب الإجابات الصحيحة لدرجة النجاح وذلك بالقسمة البسيطة على العدد الكلي للفقرات في الاختبار.

الأحكام المعتمدة على أداء المفحوصين:

ينادي العديد من أنصار الاختبارات محكية المرجع بإعداد معايير للاختبار اعتماداً على المعرفة المبنية على الأداء وذلك أثناء محاولاتهم في تطبيق الاختبار، لذلك لا بد من اختيار المحكمين الذين سيشاركون في إعداد المعايير بدقة وذلك لأن لديهم الخبرة لما سيكون عليه أداء المفحوصين النموذجي على مثل هذا الاختبار. على سبيل المثال ففي إعداد معايير الأداء لاختبار في الجغرافية الجوية في المدرسة الثانوية نفضل استخدام معلمي المدارس الثانوية كمحكمين بدلاً من اساتذة الجامعات في الطبوغرافيا. لتمييز هذا فقد اشار شبرد (Shepard, 1979) إلى ان معايير المحكمين هذه تتأثر بشكل ملحوظ بادراكهم لمعرفة كيفية اداء المفحوصين على الاختبار. ومن الأنسب درجة أكثر استخدام بيانات حقيقية لعينة مفحوصين مختارة بشكل مناسب بدلاً من الاستناد إلى محكمين مناسبين معتمدين بدرجة أكبر على انطباع خاص لمدركاتهم حول قدرات المفحوصين في الأداء على اختبار معين.

وأسلوب مبسط يستخدم البيانات المعيارية هو تطبيق الاختبار على مجموعة مفحوصين قدرتهم أقل من قدرة الفئة التي يستهدفها الاختبار، ومنها إعداد معايير لأقل كفاءة استناداً إلى وسط أداء المجموعة أو وسيطها على سبيل المثال، وصف جلاس (Glass, 1978) إعداد معايير أدنى كفاءة لامتحان شهادة الدراسة الثانوية مستنداً على وسيد أداء طلبة الصف التاسع. ومثال آخر لهذا الأسلوب هو الاختبار المعياري المستخدم في غربلة الأطفال ذوي صعوبات اللغة والذي يحدد درجة قطع لكل مستوى صفي بحيث تصلى بمقدار انحراف معياري واحد عن متوسط طلبة المستوى الصفي نفسه.

وأسلوب آخر يستخدم بيانات مجموعات مفحوصين مختلفين اختلافاً واضحاً بمستويات الكفاءة على المادة التي سيختبرون بها، وتحدد درجة القطع بحيث تعظم التمييز بين هاتين

المجموعتين وواصف بيرك (Berk, 1976) منهجية تستخدم مجموعات مفحوصين احداها درست المادة التعليمية . والأخرى لم تدرسها ووصف نبدلسكي (Nedlsky, 1954).

منهجية مجموعات المقارنة الإعداد المعايير والتي تتألف من ستة خطوات وعلى النحو الآتي:

1/ اختيار محكمين مؤهلين لديهم مألوفية بمجتمع المفحوصين

2/ السماح للمحكمين بمناقشة – إن أمكن– والموافقة على مكونات أداء ذوي أقل كفاءة مطلوبة.

3/ استخدام الأحكام لتحديد المفحوصين الكفؤين وغير الكفؤين، واستثناء أي واحد يظهر في منطقة الوسط.

4/ اختبار مجموعتي المفحوصين.

5/ رسم توزيع درجات المفحوصين لكلا المجموعتين على المتصل نفسه.

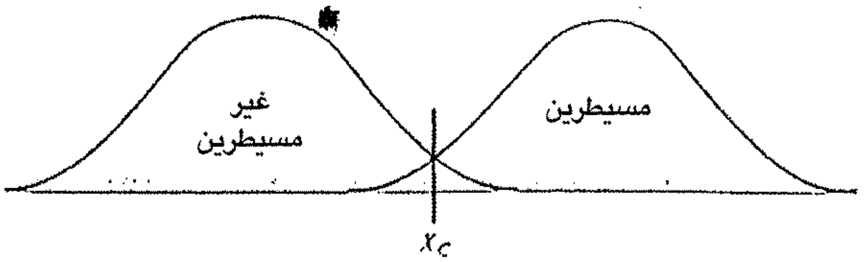
6/ إعداد معيار الأداء عند نقطة تقاطع منحنى التوزيعين (أنظر شكل (18–1)).

وطرائق أخرى لإعداد المعايير لبيانات مجموعتي مقارنة وصفها كلوفر(Kloffer, 1980) وزيكي وليفنجستون (Zieky, Livingston, 1977).

وطريقة أخرى، اقترحت من قبل نيدلسكي تسمى منهجية مجموعة الخط الفاصل (Borderline group method) وفيها يتم اختيار المحكمين وإعطاء المعلومات كما في الخطة (2،1) اعلاه، ولكن يسأل المحكمين فيما بعد لتحديد المفحوصين الذين يظهروا في منطقة الخط الفاصل في الكفاءة . وتختبر هذه المجموعة بعد ذلك ويستخدم وسيط توزيع درجاتهم في تحديد نقطة مستوى أقل كفاءة مطلوبة.

ويشير نقد هذه الأساليب إلى أن المعايير المعدة بناءاً على أداء المفحوصين تتناقض مع الهدف الأساسي للاختبارات محكية المرجع لأنه لا يوجد أية إشارة لمحتوى الاختبار في علامة أو دلالة معايير الأداء. وأوصى جلاس (Glass, 1978) بالإضافة إلى أن استخدام مجموعة درست المادة التعليمية والأخرى لم تدرسها. يبدو أنه محاولة ضعيفة لتعويض وقت الدارسة لاثبات الكفاءة. ونقد مشابه يطبق عندما تتألف المجموعات المحكية من اعضاء ممارسين للتخصص أو الوظيفة وطلبة طموحين يبحثون عن الحق في ممارسة الوظيفة أو التخصص. وفيما لو أعد معيار يميز بين مثل هاتين المجموعتين، فإن هذا يعد افتراض واضح لصدق شهادة الخبرة السابقة والتي تستدعي منطقياً مسألة الحاجة لاختبارات حالية وضعت المعايير من أجلها ومع أن هذه الانتقادات تستحق أن تؤخذ بعين الاعتبار عند تأسيس المعايير على أساس وحيد لأداء المفحوصين ويوجد دعم أكثر لاستخدام بيانات الأداء في عملية إعداد

المعايير كجزء من المعلومات المدعمة والتي يمكن استخدامها في مناقشة نتائج معايير الأداء التي تم التوصل إليها من خلال منهجيات وطرائق أخرى .



شكل (18-1) اعداد درجة قطع عند تقاطع منحنيات التوزيع التكراري لمجموعتي المقارنة

افترض على سبيل المثال أن معيار 90% صح ثم اقتراحه لمستوى الدخول في تخصص الصحة ، ولكن أشار الاختبار الميداني التمهيدي إلى أن أقل من 15% من الخريجين الحديثين في برنامج البكالوريوس لهذا التخصص يمكن أن ينطبق عليهم هذا المعيار. وتوحي هذه المعلومات بأن رخصة الممارسة تواجه معضلة جدية، فقد تكون البرامج الأكاديمية الحالية لا تعد الأغلبية العظمى من طلبتها للحياة العملية الوظيفية، أو أن عملية بناء الاختبار وصدق المحتوى النهائي معرض للمساءلة، أو أن المعيار الذي أعد كان غير منطقي. وتستحق الاحتمالات الثلاثة المذكورة جميعها أن تؤخذ بعين الاعتبار، وذلك قبل تأسس البرنامج الاختباري والتوصية بمعايير الكفاءة.

اقترح جايجر (Jaeger, 1982) عملية تجمع ما بين ظواهر الحكم الشامل على أداء المفحوصين وطريقة الحكم على الفقرات ويطلق على هذه الطريقة اسم عملية التحكيم المركب المتكرر على الفقرة، وتم إنشاء تطبيقاتها في تأسس معايير اختبارات كفاءة المدرسة الثانوية في كارولينا.

وقد استخدمت هذه الطريقة ثلاث مجموعات من المحكمين: معلمو المدارس الثانوية ومدراء المدارس والمرشدين ومواطنين في الولاية وسئل كل محكم لاكمال لاختبار وتدريج كل فقرة على مقياس نعم – لا ليجيب على لاسؤال:

" هل نستطيع كل منتظم في الدراسات الجامعية الأولية في شمال كارولينا أن يجيب على هذه الفقرة إجابة صحيحة؟" اعتقد جايجر أن هذه الكلمات اللطيفة أظهرت الحاجة لتعريف

البناء مثل ادنى كفاءة لازمة والسماح للمحكم ليركز على الأفعال الأكثر قابلية للملاحظة للحصول على الدبلوما.

ويتم تحديد درجة النجاح الموصى بها من قبل كل محكم بالجمع عبر الفقرات (بإعطاء درجة واحدة لكل إجابة نعم وصفر لكل إجابة لا) ويحدد توزيع درجة النجاح الموصى بها لكل مجموعة، وكذلك تم تسجيل توزيع استجابات نعم - لا لكل مجموعة وعلى كل فقرة تم تسجيلها. ومن ثم يشخص المحكمين الأداء الفعلي لعينة مفحوصين في الصف الحادي عشر على الفقرة (قيمة p)، وكيفية تقدير المحكمين الآخرين للفقرة نفسها. ويسمح للمحكمين بتعديل أحكامهم الأصلية في ضوء هذه المعلومات وفي دورتين متتابعتين. ويستخدم وسيط القيم لتوزيع درجات النجاح الموصى بها من قبل كل مجموعة كمعيار للمجموعة نفسها، مع أنه من المتوقع أن التجمع الأكبر للمعايير الموصى بها قد تظهر بالدورات المتتابعة. وهذا ما لم يظهر بالفعل في دراسة جايجر. ومن الواضح أن المحكمين المختلفين يستخدمون المعلومات الإضافية بطرائق مختلفة عندما يعيدون تقويم الفقرات.

الأبحاث التجريبية في منهجيات إعداد المعايير:

من خلال الأساليب المتعددة، من الطبيعي أن يطرح جماعة القياس اسئلة من مثل: هل تؤدي الطرائق المختلفة في إعداد المعايير لنتائج ملحوظة مختلفة؟ وهل تعد مجموعات المحكمين المختلفين معايير متشابهة باستخدام التقنية نفسها؟ وما هي العوامل المصاحبة أو المرافقة للتباين في إعداد المعايير من قبل محكمين مختلفين؟ لغاية الآن، أجريت دراسات تجريبية عدة للإجابة عن هذه الأسئلة، ولكن لم نصل إلى الإجابات المحددة. وسنناقش هنا بعض النتائج التوضيحية للاثبات.

دعنا أولاً نجري مقارنة عبر طرائق أحكام الفقرة، ففي إحدى الدراسات وجد أندروز وهيشت؟ (Andrews & Hecht 1971) أن محكمين ثمانية (كمجموعة) استخدموا طريقة ايبل وأعدوا معايير الكفاءة الأقل المطلوبة عند 68% إجابات صحيحة، ولكن معيار 49% إجابات صحيحة للمحكمين أنفسهم حصلوا عليها باستخدام طريقة نيدلسكي.

ولاحظ جلاس (Glass , 1978) من خلال النتائج غير المعروفة أن 95% من المفحوصين حققوا محك نيدلسكي، ولكن 50% فقط تطابقوا مع محك ايبل. ودراسة أكثر حداثة لبيونياك وزملاؤه (Behuniak , eta 1982) أجروا خلالها مقارنة بين طريقتي أنغوف ونيدلسكي في إعداد معايير اختبار تحصيلي محكي المرجع في القراءة والرياضيات وأسس محكمين عددهم ستة لاختبار القراءة متوسط نسبة إجابة صحيحة 56.7% كمعيار بطريقة انغوف و 43.2% بطريقة نيدلسكي.

وأما في اختبار الرياضيات فكانت النتائج عكس ذلك، وإذ كانت النسبة 70.2% بطريقة انغوف و 77.1% بطريقة نيدلسكي . ومع أن دراسات المقارنة هذه لا توفر أدلة لأفضلية طريقة على غيرها، إلا أنها أثبتت بوضوح أن طرائق تحكيم الفقرة المختلفة لا يمكن عدّها توائم في إعداد المعايير . وكانت النتائج غير متطابقة في الدراسة الأخيرة وذلك لأن مجموعة المحكمين قسمت إلى نصفين، وهنالك فروق أساسية بين المعايير الموصى بها من قبل مجموعات المحكمين المختلفة باستخدام الطريقة نفسها.

وقارن كل من كوفلر (Koffler, 1980) وميلز (Mills , 1983) نتائج إعداد المعايير لطرائق الحكم على محتوى الفقرات والطرائق التي تستخدم بيانات المفحوصين. وعلى وجه الخصوص فحص كوفلر إعداد المعايير بطريقة نيدلسكي وأسلوب مجموعات المقارنة بينما استخدم ميلز اسلوب انغوف لمجموعات المقارنة وتقنياتمجموعة الخط الفاصل، ومع ذلك فقد كانت النتائج مختلطة (ممزوجة) نوعاً ما. فقد أظهرت دراسة ميلز أنه عند استخدام ثلاثة طرائق أو أكثر فمن الممكن الحصول على اتفاق بين طريقتين منهما على الأقل.

فعلى سبيل المثال كانت معايير طريقة انغوف وطريقة مجموعات المقارنة أكثر تطابقاً مما هي مع معايير طريقة مجموعة الخط الفاصل في معظم الحالات المدونة في الدراسات ولتوضيح التنوع في المعايير التي قد تنتج في الطرائق المختلفة، يبين جدول (18–2أ) بعض نتائج دراسة ميلز، إذ بدون الاختبارات معينة قائمة معايير النسبة الصحيحة، كذلك يبين جدول (18–2ب) نسبة المفحوصين الذين حققوا كل معيار.

وبينت دراسات أخرى تجريبية لإعداد المعايير أن المحكمين بخصائصهم المختلفة قد يصلوا إلى توصيات مختلفة تماماً. فعلى سبيل المثال دون جايجر (Jager,1982) فوارق مهمة في المعايير الناتجة عن مجموعات مؤلفة من معلمين ومواطنين عندما راجعوا فقرات اختبارات كفاءة المدرسة الثانوية شمال كاليفورنيا إضافة إلى أن ساوندرز وزملاؤه (Saunders, etal, 1981) وجدوا علاقة بين مستوى معرفة المحكمين للمحتوى ودرجة النجاح التي حصل عليها من المحكمين بطريقة نيدلسكي ومن مراجعة مثل هذه النتائج، ولاي دراسة إعداد معايير من الأنسب الحساب التجريبي لمدى مساهمة كل من الفروق بين الأحكام وطرائق إعداد المعايير النهائية الموصى بها. وقد اكتشف كل من بريتان ولوك دور (Brenan & Look-wood, 1980) قابلية نظرية إمكانية التعميم للتطبيق في هذا السياق، كذلك فإن هذا المجال يحتاج للمزيد من البحث والعمل.

جدول (18–2أ) : أدنى معايير نجاح (من خلال النسبة الصحيحة) المحصلة من طرائق مختلفة من دراسة دونها ميلز 1983.

صيغة الاختبار	طريقة انغوف	مجموعة المقاونة	مجموعة الخط الفاصل
G	0.68	0.68	0.87
H	0.68	.75	0.88
I	0.62	.65	0.80
J	0.78	.62	0.90
K	0.68	.70	0.85
L	0.68	.60	0.83

جدول (18–2ب) : نسب المفحصوين الذين فشلوا في النجاح في المعايير الناتجة عن طرائق مختلفة.

صيغة الاختبار	طريقة انغوف	مجموعة المقاونة	مجموعة الخط الفاصل
G	7.06	6.30	29.75
H	6.03	8.19	22.09
I	7.60	9.42	26.66
J	9.10	4.07	23.71
K	7.82	9.32	25.40
L	7.95	5.38	23.11

أخيراً، فبالإضافة إلى احتمالية الحصول على نتائج مختلفة من قبل عينات محكمين مختلفة، وتقنيات مختلفة ، فقد أشار فان دير لندن (Vander Linden, 1982) أن عدم توافق الأحكام الداخلية قد يكون مشكلة ويظهر هذا عندما يؤشر محكم احتمالية منخفضة لمفحوص يمتلك أقل كفاءة مقبولة للنجاح في اختبار فقرات سهلة، واحتمالية نجاح المفحوص نفسه في اختبار فقراته صعبة، فقد اقترح معامل تناقض يعتمد على نظرية الاستجابة للفقرة، وأسس تطبيقاتها مع الأحكام التي حصل عليها من طريقتي نيدلسكي وانغوف. وعند الأخذ بعين الاعتبار الحاجة إلى بذل جهود في هذا المجال، اقترح شيبرد (Shepard, 1980) أبحاث تطبيقية إضافية لإعداد المعايير يجب أن تسوّغ لبرامج اختبارات التأهيل غير المتميزة. وللتعلم أكثر

عن أثر التعليمات المختلفة المقدمة للمحكمين . ومزايا وعيوب عمل المحكمين المستقل أوضحت مجموعات، وكذلك الوقت الأكثر مناسبة والصيغ التي يفضل بوساطتها عرض البيانات المعيارية على المحكمين.

اعتبارات عملية في إعداد المعايير:

بغض النظر عن الطريقة المستخدمة فإنه لا يمكن الاستغناء عن الحاجة إلى أحكام في تأسيس معايير الأداء الاختباري.

ومن المهم تمييز أن إعداد المعايير عبارة عن مشكلة سيكومترية مهمة. فهي ليست قضية تقنية فحسب، بل أن توابع ملائمة المعايير أو عدمه للأفراد والمعاهدة والمجتمع يجب اخذها بعين الاعتبار. وللبعض مثل جلاس (Glass, 1972) فإن النتائج. النظرية والأسس التجريبية للمعرفة التي تخضع لها أساليب إعداد المعايير إن كانت غير ملائمة فإن التطبيق نفسه يبدو غير قابل للتعديل ولاخرين (مثل popham, 1978, Hambelton, Hambelton 1978, 1980 (Scriven, 1978, & Shepard, 1976, 1979.

فإن الحاجة إلى المعايير موجه لاتخاذ قرارات في سياقات تربوية وتحليليه، وهذا واقع لا يمكن الغاؤه او إهماله حتى يتم حل المشكلات الفلسفية والمنهجية المتعلقة بالتطبيقات العملية. وباعطاء هذه الحالة من المضاهاة وصولاً إلى مشكلة اعداد المعايير، فإن الخطوة الأولى هي الاستفسار عما إذا كانت هنالك حاجة ملحة ومنطقية لتأسيس معايير أداء لتفسير درجات الاختبار لا زالت قيد المساءلة . فعلى سبيل المثال ففي مجال مثل الدراسات الاجتماعية في المدارس الابتدائية فليس من الضروري أن يسيطر الطلبة سيطرة تامة على المادة في وحدة ما قبل تدريس وحدة أخرى. ومع أن تقييم تحصيل الطلبة يكون مناسباً إلا أن الحاجة إلى تحديد درجات قطع كمؤشر لمستويات السيطرة لمثل هذا الاختبار هو موضوع مطروح للمساءلة . في حين أن الحاجة واضحة لمعايير أداء لتفسير درجات اختبار مستخدم في اتخاذ قرار. وينصح هنا بمراجعة الأساليب الشائعة لإعداد المعايير وتحديد تلك التي يبدو أنه يمكن الدفاع عنها في الموقف المعطى.

بعد ذلك، فمن المهم تحديد المهددات المشابهة لعدم صدق الاستنتاجات التي أخذت من درجات الاختبار. وقد حدد جايجر (Jaeger, 1979) عدداً من مهددات صدق الاستنتاجات المأخوذة في درجات الاختبار إلى درجات النطاق. وربط المشكلات المختلفة بطريقة معينة في إعداد المعايير. وتلك التي تطبق على طرائق التحكيم جميعها والتي تعتمد على محتوى الفقرة تكون متميزة، وذلك لأسباب منها التحديد غير المناسب للنطاق، والخطأ العشوائي المؤثر على قرارات المحكمين المختلفة، والطرائق غير المناسبة أو غير الملائمة لمعاينة الفقرات للتأكد من تمثيلها للنطاق وتحيز الأحكام في مراجعة الفقرات منفردة . لاحظ أن المهددات الأولى والثالثة

والرابعة ظهرت بالفعل في طرائق بناء الاختبار الخاطئة (أو على الأقل غير ملائمة لتؤكد الدرجة اللازمة في صدق المحتوى لأنواع الاستنتاجات المرغوب بها بوساطة درجات الاختبار) ومن المحتمل أن التهديد الثاني والخامس التي يمكن ضبطها أو خفضها، أثناء إعداد المعايير، بافتراض أن المعايير تأسست لدرجات طوّر اختبارها مسبقاً. لذا يبدو أن استخدام عدد كبير من المحكمين (ويفضل أكثر من 6 إلى 8 الذين استخدموا في دراسات منشورة سابقاً) واختيارهم عشوائياً من مجموعة محددة بعناية من المحكمين المؤهلين وتزويدهم بمعلومات واضحة عن السياق التي ستستخدم فيه درجات الاختبار، وتدريبهم على أداء المهمة لتقليل المشكلات المحتملة. وقد زودنا كل من زيكي وليفنجستون (Zieky & Livingston, 1977) وهاملتون (Hambelton, 1978) بإرشادات إضافية ومقترحات تحقق طرائق عدة في إعداد المعايير التي نوقشت هنا.

إضافة إلى ذلك ، ينصح باستخدام اثنين أو أكثر من أساليب إعداد المعايير وعينات محكمين عديدة تختار عند الضرورة لتمثيل وحدات محكمين مختلفة وملائمة. وفيما لو فشلت الطرائق المتنوعة في إنتاج معيار مفرد موصم (كما يجب أن تظهر) فإن السؤال عن تلك التي يجب اختيارها أو كيفية المواءمة لتكوين حل مركب مناسب فهذا يتطلب في النهاية حكماً قيمياً.

وينصح أيضاً فحص الأدلة التجريبية حول كيفية اداء المفحوصين النموذجي في الاختبار واستخدام هذه المعلومات في تقويم نتائج وضح معايير معينة. افترضعلى سيبل المثال حالة تأسس الاداء فيها باحدى طرائق تحكيم الفقرات عن درجة قطع 87% عن اختبار تعلم للاتقان في الرياضيات لأطفال مرحلة الطفولة المتاخرة. ولكن نتائج التجريب الميداني بينت أن نسبة بسيطة من الطلبة قد حققوا هذا المحكم بعد التعليم. أن نتائج وضع هذا المعيار يجب أن تؤخذ بعين الاعتبار من خلال عدة وجهات نظر. الأول، أنه يجب أن يؤخذ بعين الاعتبار توابع التعلم المتتابع، فإن تقدم الأطفال دون أن يحققوا هذا المعيار فماذا سيكون تأثير هذا على التعلم المستقبلي ؟ بالإضافة إلى أن أثر الدافعية من مواجهة إعادة التعليم يجب أن تؤخذ بعين الاعتبار. والفرض الضائعة (من خلال تقليل الوقت لتعليم الموضوعات الأخرى) قد يكون عاملاً مشابهاً في مجالات مثل شهادة التخصص، فإن نتائج (عواقب) تأهيل المتقدمين الذين هم أقل من أدنى كفاءة مطلوبة يجب أن توازن مقابل خسارة المجتمع للخدمات التي يقدمها أصحاب الشهادات المؤهلين تأهيلاً حقيقياً، والذين فشلوا في تحقيق معيار متشدد. ومن الواضح أن تأسيس معايير أداء مفيدة تتطلب موازنة جيدة لتقنية ملائمة ورؤية واضحة لأهداف وغايات البرنامج الذي يتطلب قرارات عن المفحوصين تعتمد على أدائهم في الاختبار. إحدى الأساليب تسمح لمعد المعايير أن يأخذ هذه التوابع بانتظام تتطلب تطبيق نظرية القرار، والتي ستناقش في الأجزاء الآتية.

اعتبارات تقنية في إعداد المعايير:

يركز الأدب التقني في إعداد المعايير على إحدى القضايا الأساسية، هي إعداد درجة قطع على التدريج الملاحظ تسمح لمستخدم الاختبار تكوين خلاصات عن حالة السيطرة الحقيقية للمفحوصين والتي تم تحديدها على متصل درجات النطاق. وكما سنرى، فإن المعنى الملائم للخلاصات المناسبة يعتمد على الطريقة المستخدمة في إعداد درجة القطع على التدريج الملاحظ . وفي بعض الطرائق تحدد الملائمة من خلال احتمالية التصنيف الخاطئ. وستصف في هذا الجزء أساليب عدة لهذه المشكلة، وكل أسلوب موصوف هنا هو تطبيق لنظرية القرار والتي تم عرضها في الفصل الثاني عشر. كذلك فإن الأدب التقـني طرح قضية تقنية أخرى ثانوية نوعاً ما، ويعتمد تطوير معامل لتكميم نوعية القـرارات المأخوذة على درجة القـطع على التدريج الملاحظ. والأبحاث التي تعالج هذه المشكلة تجـدها في (Mellen berg, 1977) (Wil- cox, 1977) و (Hunyh, 1980) يمكن للقارئ المهتم بالموضوع الرجوع إليها.

إعداد درجة قطع على التدريج الملاحظ:

تقليل احتمالية التصنيف الخاطئ: تقسم درجة قطع تدريج النطاق درجة النطاق إلى منطقتين: مفحوصين عند درجة القطع أو أعلى منها وهؤلاء مسيطرين حقيقيين، ومفحوصين درجاتهم أدنى من درجة القطع وهؤلاء غير مسيطرين حقيقيين وتفسر درجة النطاق للمفحوص على أنها نسبة الفقرات من النطاق التي يستطيع المفحوص أن يجيب عنها إجابة صحيحة، كذلك فإن مدى تدريج درجة النطاق يمتد من صفر إلى 1.

وسنرمز إلى درجات النطاق بالرمز t، ودرجة القطع على تدريج النطاق وكما لاحظنا من البداية أن t_0 قد تستخدم في أي من طرائق إعداد المعايير الموصوفة في الأجزاء السابقة. وبما أن تدريج درجة النطاق هو تدريج لدرجات النسبة الصحيحة وطرائق إعداد المعايير ينتج عنها درجات إما على شكل نسبة مئوية أو تدريج عدد الإجابات الصحيحة فمن الضروري تحويل درجات القطع هذه قبل تفسيرها على أنها t_0 . وفيما لو تم التعبير عن درجة القطع على متصل النسبة الصحيحة فإنه يجب تقسيمها على 100 للحصول على t_0 : وفيما لو تم التعبير عنها على تدريج العدد الصحيح فيجب تقسيمها على عدد الفقرات المستخدمة في إعداد درجة القطع من أجل الحصول على t_0.

وحالما تتحدد t_0، فإن المشكلة تكمن في تحديد درجة القطع على تدريج الدرجات الملاحظة والتي تسمح لمستخدم الاختبار عمل استنتاجات مناسبة حول حالات السيطرة للمفحوصين. وعادة يرمز لدرجة القطع الجديدة بالرمز X_0 . ويعبر عنها على تدريج العدد الصحيح والذي يرمز له بالرمز X. وحالما تختار قيمة X_0، فإن كل مفحوص يكون في إحدى الفئات الآتية:

1) صح إيجابي: حصل على درجة نطاق تساوي أو أكبر من t_0، ودرجته الملاحظة تساوي X0 أو أكبر منها .

2) صح سلبي : حصل على درجة نطاق أقل من من t_0 ودرجة ملاحظة أقل من X0.

3) خطأ إيجابي: حصل على درجة نطاق أقل من t_0 ودرجة ملاحظة تساوي أو أكبر من X0.

4) خطأ سلبي: حصل على درجة نطاق تساوي t_0 أو أقل منها ودرجة ملاحظة أقل من X0.

إحدى طرائق حل هذه المشكلة يتضمن اختبار قيمة X0 التي تقلل احتمالية التصنيف الخاطئ للمفحوصين (سواء الخطأ الإيجابي أو السلبي) والمشكلة التي تواجه تحقيق (هذه الإستراتيجية هو أننا لا نعرف بالفعل احتمالات كل من الخطأ الإيجابي والخطأ السلبي. وللتغلب على هذه المشكلة نستخدم تقديرات هذه الاحتمالات، ويتطلب تقدير هذه الاحتمالات عمل افتراضات تتعلق بالتوزيع ثنائـــي المتغير لدرجات النطاق والدرجات الملاحظة.

والافتراض النموذجي لتوزيع ثنائي المتغير لكل من درجات النطاق والدرجات الملاحظة هو التوزيع من نوع بيتا – ثنائي (β- binomial dist) واستخدم هونيه **(Hunyh, 1976)** هذا الافتراض في اشتقاق نتائج أدت إلى تحديد X0 (C0 برموز هونية) التي قللت احتمالية التصنيف الخاطئ، وطريقة هونيه أنه إذا كان عدد الفقرات كبيراً و t_0 ليست قريبة من الصفر أو الواحد فإنه يمكن تقريب X0 على النحو الآتي:

$$X_0 = \frac{KR/21 \ldots n}{K\ R(21)} + \frac{1-(KR21)}{K\ R\ (21)}\ \mu_M + \qquad (2\text{-}18)$$

حيث تمثل n عدد فقرات الاختبار. وتظهر دقة هذا التقريب إذا تألف الاختبار من 20 فقرة أو أكثر، وكانت t_0 في المدى بين 0.50 إلى 0.80. **(Hunyh, & Saunders,1980)**. ويبين جدول (18–2) درجات قطع تقريبية ناجمة عن تطبيق المعادلة (18–3) على ستة مجموعات مفحوصين مختلفة. ويمكن أن يظهر نمط نتائج مشابه فيما لو استخدمت طريقة هونيه نفسها بدلاً من الطريقة التقريبية. كذلك قد يظهر نمط مشابه من النتائج في حالة توافر يتم KR 20 ومتوسط الدرجات الملاحظة للمجتمع بدلاً من تقديرات العينة لهذه الكميات ولكل مجموعة من

المجموعات الستة N=28 , t_o = 0.75، ومع ذلك فإنها تختلف في KR21 و / أو $\hat{\mu}_M$ وبتفحص السطر الأفقي KR21= 0.6 أو KR21= 0.8 نرى إن X0 تعتمد على متوسط درجة المجموعة. وبازدياد $\hat{\mu}_x$ فإن X_0 تقل. ولأخذ هذه النتائج وفقاً لعلاقاتها الصحيحة، افترض أن t_o كانت 0.75 و N = 0.28 وبالحدس. علينا أن نتوقع أن X_0 تكون n_{t0}= 21، وهو أننا نتوقع أن درجة القطع على التدريج الملاحظ تتطلب من المفحوص أن يجيب على 75% من الفقرات إجابة صحيحة لنقرر أنه مسيطر. وبتمحيص أي من السطور نرى أن X0 تساوي 21 تقريباً عند $\hat{\mu}_x$=0.21 ، ومع ذلك فإن X0 تتجاوز 21 عندما تكون $\hat{\mu}_x$ أقل من 21، وتكون أقل من 21 عندما تكون $\hat{\mu}_x$ أكبر من 21. لذا فإن المفحوصين من مجموعة متوسطها أكبر من n_{t0} تكون درجة القطع لهؤلاء على متصل الدرجة الملاحظة أقل تشدداً من درجة القطع لمفحوصين من مجموعة لها متوسط أقل من t_o.

يوجد على الأقل تفسيرين حدسيين لاعتماد X0 على $\hat{\mu}_x$ احدها، عندما تكون درجات معظم المفحوصين أعلى من n_{t0} فإن معظم هؤلاء يكونوا مسيطرين وطريقة وضع X0 التي تعكس هذا الاعتقاد هو وضع درجة قطع ملاحظة أقل. والتفسير الثاني مرتبط بتقديرات الانحدار للدرجات الحقيقية تذكران معادلة تقدير الدرجة الحقيقية هي:

$$(\hat{\mu}_x - X) \quad \rho_{xx}\backslash + \hat{\mu}_x = \hat{T}$$

وكما تبين هذه المعادلة، فإنه يتوقع أن تكون الدرجات الحقيقية للمفحصوين الذين درجاتهم الملاحظة اقل من المتوسط أعلى من درجاتهم الملاحظة، أما المفحوصين الذين درجاتهم الملاحظة أعلى من المتوسط فيتوقع أن تكون درجاتهم الحقيقية أقل من درجاتهم الملاحظة. ماذا يمكن أن يحدث إذا كان لدينا مفحوصين أثنين لهم $\hat{\mu}_M$ =21، وجاء أحدهم من مجموعة لها $\hat{\mu}_M$ =17، والأخر من مجموعة لها $\hat{\mu}_M$ = 0.25 إذا كانت KR21 تساوي 0.6، فإن الدرجة الحقيقية المحسوبة للمفحوص الأول = 19.4 وللآخر= 22.6 (أو 21) لذا فإن كانت n_{t0} تستخدم على أنها درجة القطع على التدريج الملاحظ فإن المفحوص الأول سيتم تصنيفه. على أنه غير مسيطر والآخر مسيطر. وبدلاً من استخدام الدرجات الحقيقية المقدرة فإن طريقة هونية تعدل X0 لتعكس العلاقة بين الدرجات الملاحظة ودرجات النطاق.

جدول (18–3) : أثر التغير في KR21 و $\hat{\mu}_x$ على Xo

μ_x			
25	21	17	**KR21**
18.10	20.76	23.43	0.6
19.25	20.75	21.75	0.8

n = 28 ، t_0 = 0.75 لكلا المجموعتين

ويبين جدول (18–3) أنه بازدياد KR21 فإن تباين X_0 دالة لنقصان $\hat{\mu}_x$. إلا أنه كلما اقتربت KR21 من 1.0 فإن $\hat{X}_0$ تقترب من n_{t_0} + 0.5 ، وبغض النظر عن قيمة $\hat{\mu}_x$، فإن خطأ القياس يقل وكذلك قيمة $\hat{X}_0$ المعدلة تقل أيضاً.

وتظهر مشكلة عملية عند استخدام طرائق مثل طريقة هونية في البرامج الاختبارية التي تطبق باستمرار. للتوضيح ، تخيل أن برنامج اختياري سيستخدم فيه الاختبار نفسه في ثلاثة تطبيقات اختبارية يفصل بين كل تطبيقين سنة. ويتألف الاختبار من (100) فقرة و t_0= 0.70 ، وفي السنة الأولى كانت KR21 = 0.75 و $\hat{\mu}_x$= 0.7 لذا فإن درجة القطع X_0 تقرب على النحو الآتي:

$$69.46 = 0.5 + (71)\ \frac{1-.75}{75} + (0.70)\ \frac{0.75 - 100}{0.75}$$

أو 69 لأقرب عدد صحيح. افترض أنه في السـنة الثانية كانت KR21= 0.75 ثابتة ولكن $\hat{\mu}_x$ = 0.74 لذا فإن التقريب الجديد لدرجة القطع X_0 = 68 لأقرب عدد صحيح. والسؤال الذي يواجه مستخدم الاختبار في السنة الثانية يتعلق بأي من درجات القطع سيستخدم، هل 69 من السنة الأولى أم 68 من السنة الثانية. لاحظ أن متوسط الدرجات في السنتين كان متشابهاً، وهذا دليل على أن كلا العينتين تمثل المجتمع نفسه. لذا فإن التوزيع ثنائي الحد لدرجات النطاق الملاحظة هي نفسها في كلا السنتين، وهكذا فإن (69، 68) تعد تقديرات لدرجة القطع نفسها. وينصح مستخدم الاختبار ثانية باستخدام درجة القطع 69 وذلك للحفاظ على الاتساق من سنة لأخرى. والآن افترض أنه كانت قيمة KR21 في السنة الثالثة 0.75 ولكن $\hat{\mu}_x$ = 81 والتقدير الجدير لدرجة القطع يكون 66 وثابتة. نطرح السؤال نفسه. أي من درجات القطع نستخدم، إذ أن متوسط درجات السنة الثالثة يختلف كثيراً عن

متوسط درجات السنتين السابقتين. وهذا مؤشر إلى أن عينة السنة الثالثة وعينات السنتين الأولى والثانية هي عينات لمجتمعات مختلفة. وبما أن درجة قطع الدرجات الملاحظة يعتمد على مجتمع المفحوصين فإن درجة القطع الجديدة (66) ودرجات القطع القديمة (68 و 69) من المحتمل أنها تقدير لدرجات قطع مختلفة، لذا لتقليل احتمالية التصنيف الخاطئ لمفحوصي السنة الثالثة فإنه من الأنسب المستخدم $X_0 = 0.66$ ولكن استخدام درجة القطع المحسوبة هذه تسبب بالتأكيد مشكلات قانونية (أخلاقية)، إذن فإن مفحوصي السنة الثالثة بمتوسط 66 سينجحون، ولكن مفحوصي السنتين الأولى والثانية الذين حصلوا على 66 سيفشلون في الاختبار، وهنا يظهر تناقض بين المبادئ السيكومترية والأخلاقية.

وهنالك أسلوب لإعداد درجة قطع للتدريج الملاحظ ويلغي هذه المشكلة، وفضلاً عن إعداد t_0، فمن الضروري إعداد كل من :

t_1 : وهي أعلى درجة نطاق تقع تحديداً في منطقة عدم السيطرة.

t_2 : وهي أدنى درجة نطاق تقع تحديداً في منطقة السيطرة.

ويطلق على المنطقة الواقعة بين t_1 و t_2 نطاق عدم الاختلاف (Ind: fferene zone) ، ونحن غير متأكدين من أن المفحوص الذي تقع درجته في هذا المدى مسيطراً ، وكنتيجة لا نهتم باحتمالية التصنيف الخاطئ للمفحوص الذي تقع درجته في هذا المدى، ولكننا نهتم باحتمالية التصنيف الخاطئ للمفحوص الذي درجة نطاقه t_1 على أنه مسيطر، وكذلك باحتمالية التصنيف الخاطئ للمفحوص الذي درجة نطاقه t_2 على أنه غير مسيطر. ونتيجة لذلك فإننا نحدد X_0 بحيث تكون أكبر هذه الاحتمالات أصغر ما يمكن. وتسمى هذه الطريقة تصغير القيمة الكبرى (mininax, procedure) (Hunyh, 1980b) وذلك لأنها تقل أكبر احتمالية للتصنيف الخاطئ. ومن جهة أخرى فإن X_0 المحددة بهذه الطريقة لا تتغير بتغير مجتمع المفحوصين (وهناك طرائق أخرى لطريقة تصغير القيمة الكبرى تتغير فيها قيمة X_0 بتغير مجتمع المفحوصين).

ولتطبيق هذه الطريقة تحسب احتمالية كل من التصنيف الخاطئ لمفحوص درجة نطاقه t_1 واحتمالية التصنيف الخاطئ لمفحوص درجة نطاقه t_2 ولكل درجة قطع محتملة. ويبين جدول (4-18) هذه الاحتمالات $t = 0.7$, $t = 0.85$ ولاختبار مؤلف من 7 فقرات. وحالما تحسب هذه الاحتمالات فإنه يتم تحديد أقصى احتمالية عند كل درجة قطع، ويؤشر لكل قيمة منها بنجمة (*) ، ويتم اختيار الدرجة المرتبطة باقل احتمالية قصوى من هذه على أنها X_0 ويتفحص هذه الاحتمالات المشار إليها بالنجمة فإننا نختار $X_0=6$. وعند درجة القطع هذه فإن أكبر احتمالية للتصنيف الخاطئ تكون لمفحوص درجة نطاقه $t_1 = 0.3295$ وعلى الرغم من أن X_0 هي درجة القطع لأصغر قيمة كبرى إلا أن المفحوصين الآخرين يكون احتمالية التصنيف الخاطئ لهم أقل من 0.3295.

ولحساب احتمالات التصنيف الخاطئ من الضروري تأسيس افتراض يتعلق بتوزيع الدرجات الملاحظة المفحوصين درجة نطاقهم t_1 ولمفحوصين درجة نطاقهم t_2 واحد من الافتراضات الشائعة أن توزيع كل منهما هو توزيع ثنائي الحد (عرض في الفصل السادس). ويمكن استخدام التوزيع ثنائي الحد كمثال في حساب احتمال عدد مرات ظهور الرأس m عند رمي قطعة نقد لمرات عددها r. ويمكن استخدام هذه الطريقة حتى إذا كانت احتمالية ظهور الرأس ليست 0.5 . والمنطق في استخدام التوزيع ثنائي الحد في إعداد المعايير هو أن درجة النطاق عبارة.عن نسبة الفقرات من النطاق التي يستطيع المفحوص الإجابة عنها إجابة صحيحة. لذا فإن درجة النطاق يمكن تفسيرها على أنها احتمالية إجابة المفحوص إجابة صحيحة عن فقرة اختيرت عشوائياً من النطاق.

ويستخدم التوزيع ثنائي الحد ف حساب احتمالية إجابة مفحوص درجة نطاقه = t إجابة صحيحة على فقرات عددها X اختيرت عشوائياً من فقرات كلية عددها n.

جدول (18-4) احتمالات التصنيف الخاطئ لدرجات نطاق 0.70 و 0.85 لدرجات القطع الملاحظة المحتملة جميعها

درجة القطع	درجة النطاق	
	*0.70	0.85
صفر	*0.000	0.0000
1	*0.9999	0.0000
2	*0.9963	0.0000
3	*0.9713	0.0011
4	*0.8741	0.0120
5	*0.6472	0.0737
6	*0.3295	0.2834
7	*0.0824	*0.6749

* تشير إلى أقصى احتمالية للتصنيف الخاطئ لكل درجة قطع

ويبين الجدول (18-5) احتمالية كل درجة ممكنة لاختبار افتراضي مؤلف من 7 فقرات لمفحوصين درجات نطاقهم $t_1 = 0.70$ و $t_2 = 0.85$ والتي حسبت باستخدام التوزيع ثنائي الحد . وتعتمد الاحتمالات في جدول (18-4) على الاحتمالات المبينة في جدول (18-5) ولتبيان كيف تم الحصول على النتائج في جدول (18-4) من بيانات جدول (18-5)، افترض

أن X0 وضعت عند0 ، والمفحوص الذي درجة نطاقه 0.7 (غير مسيطر) سيكون تصنيفه خاطئ فيما لو حصل على درجة 0 أو أعلى في الاختبار. وعلى وفق جدول (18–5) فإن احتمالية حصول المفحوص الذي درجة نطاقه 0.7 على درجة 0 أو أعلى تساوي (0.3177+ 0.2471 + 0.824 =0.6472) وهذه احتمالية خطأ (الخطأ الموجب) لمفحوص درجة نطاقه 0.7 عند درجة قطع 0 لمفحوص درجة نطاقه 0.85 (مسيطر) يكون تصنيفه خاطئ إذا حصل المفحوص على درجة اختبار 4 أو أقل، وثانية وفق جدول (18-5) فإن هذه الاحتمالية تساوي (0.0167 - 0.0109 + 0.0012 + 0.001 + 0.00 = 0.737) ويمكن تحديد احتمالات التصنيف الخاطئ لدرجات القطع الأخرى بطريقة مشابهة.

تذكر أنه لاختبارنا الافتراضي المؤلف من 7 فقرات، فإن أقصى احتمالية للتصنيف الخاطئ للمفحوص تساوي (0.3295) افترض أن مستخدم الاختبار يأمل استخدام هذا الاختبار، ووجد أن هذه الاحتمالية عالية بشكل مفرط. وإن كان عملياً زيادة طول الاختبار فإنه يمكن تقليل هذه الاحتمالية. وتقترح هذه الحقيقة طريقة لتحديد طول الاختبار و X0 . أولاً: قرر وعلى ضوء أقصى احتمالية للتصنيف الخاطئ، ثم أعد جدول مشابه لجدول (18–4) للاختبار بطوله الأصلي واختر بشكل مؤقت قيمة X0. وإن كانت X0 هذه لا تعطي أقصى احتمالية والتي تكون أقل من احتمال المحك المعُد، بعدها أضف إلى الاختبار فقرة وحدد قيمة X0 الجديدة، واحتمالات التصنيف الخاطئ. استمر بهذه الطريقة حتى تصل إلى طول اختبار مناسب وكذلك X0 وقد ناقش كل من فاهز **(Fahner, 1974)** وويلكسوكس **(Wilcox, 1976)** هذه الطرائق بتفاصيل أكثر.

جدول (18–5) : توزيعات الاحتمالية لاختبار مؤلف من 7 فقرات لدرجات النطاق 0.70 , 0.85

درجة القطع	درجة النطاق	
	0.70	0.85
صفر	0.0002	0.0000
1	0.0036	0.0001
2	0.0250	0.0012
3	0.0972	0.0109
4	0.2269	0.0617
5	0.3177	0.2096
6	0.2471	0.3960
7	0.0824	0.3206

تقليل الكلفة المتوقعة للتصنيف الخاطئ:

سيتم وصف أسلوبين يهتمان بتقليل احتمالية التصنيف الخاطئ. ويكون هذا الأسلوب مناسباً عندما تكون الأخطاء : الخطأ الإيجابي. والخطأ السلبي متساوية في جديتها. وفي العديد من المواقف لا يكون هذا صحيحاً. فعلى سبيل المثال ففي المنهج المنظم هرمياً يكون خطأ – الخطأ السلبي. وعندما يتم عمل خطأ من نوع الخطأ – الإيجابي فإن التلميذ سينتقل إلى الوحدة التالية قبل أن يكون جاهزاً للانتقال، وفي حالة خطأ من نوع الخطأ – السلبي فإن التلميذ سيبقى في الوحدة نفسها. ومع أن استمرار التلميذ في الوحدة نفسها سيؤدي فقط إلى مراجعة قصيرة نسبياً للوحدة، وبالتالي لا يكون الخطأ جدياً بدرجة كبيرة.

وتستخدم الطرائق التي تأخذ خسارة التصنيف الخاطئ بالحسبان مفهوم الخسارة المتوقعة. وسنوضح هذا المفهوم وهذا النوع من الطرائق بتطبيق طريقة تصغير القيمة الكبرى لاختبار الفقرات السبع الافتراضي المستخدم في الجزء السابق. وبدلاً من تحديد درجة القطع التي تقلل أكبر احتمالية للتصنيف الخاطئ. فإن طريقة تصغير القيمة الكبرى الجديدة ستقلل أكبر خسارة ناجمة عن التصنيف الخاطئ.

افترض أن خسارة الخطأ الإيجابي 5 وخسارة الخطأ السلبي 0.5 (ولغاية الآن سنتجاهل قضية كيف تم الحصول على هذه القيم). دعنا نأخذ بعين الاعتبار توابع اختيار$X0=4$ ، ومع درجة القطع هذه فإن المفحوص ذو درجة نطاق $t_2 = 0.70$ تكون احتمالية تصنيفه خطأ (خطأ إيجابي) تساوي 0.8741 (أنظر جدول 18–4). والخسارة المتوقعة تساوي خسارة التصنيف الخاطئ مضروباً باحتمالية التصنيف الخاطئ أو $5 \times 0.8749 = 4.3705$ والخسارة المتوقعة للتصنيف الخاطئ لمفحوص درجة نطاقه $t_2 = .85$ تساوي $0.5 \times 0.210 = 6.0.$ وإجراء هذه الحسابات لدرجات القطع المحتملة جميعها تؤدي إلى النتائج المبينة في جدول (18–6). ودرجة القطع التي تؤدي إلى الأقل من قيم تصغير القيمة الكبرى عند $X_0 = 7$ وبأخذ الخسارات المتوقعة بالحسبان، فإن درجة القطع أكبر من تلك التي تم الحصول عليها بتخفيض الاحتمالية الأكبر للتصنيف الخاطئ. وهذه تظهر لأن خساراتنا تشير إلى أن خطأ الخطأ – الإيجابي يكون جدياً عشرة أضعاف ما هو للخطأ السلبي. وبالتالي فإننا نحتاج إلى درجة عالية قبل أن نستنتج أن المفحوص مفحوص مسيطر.

جدول (18-6): الخسارات المتوقعة لدرجات نطاق 0.7 و 0.85 لدرجات القطع الملاحظة المحتملة جميعها.

درجة القطع	درجة النطاق	
	0.70	0.85
صفر	*5.0000	0.00000
1	*4.9995	0.00000
2	*4.9815	0.00000
3	*4.8565	0.00055
4	*43705	0.00600
5	*3.2360	0.03685
6	*1.6475	0.14170
7	*0.4120	0.33745

* تشير إلى أقصى خسارة متوقعة لكل درجة قطع محتملة

وتقلل طريقة تصغير القيمة الكبرى أكبر خسارة متوقعة، وهذه تمكننا من تطوير طريقة تقلل الخسارة المتوقعة الكلية. وقد استخدم هونيه (Hunyh, 1976) النموذج ثنائي الحد - بيتا لتطوير مثل هذه الطريقة وثانية تطبيق النتائج معقد نوعاً ما، ولكن التقريب الدقيق متوفر للاستخدام وذلك عندما تكون $n \leq 20$، و t_0 تتراوح ما بين (0.50–0.80) وهذا التقريب هو:

$$0.5 + \hat{\mu}_x \frac{1-KR(21)}{KR(21)} + (t_o - 1)t_o \sqrt{\frac{*KR(21)-n}{KR(21)}} Z + t_o \frac{KR(21)-n}{KR(21)}$$

(3-18) ....

وفي هذا التعبير فإن Z تمثل الدرجة في التوزيع المعياري الطبيعي عندما تكون المساحة $\frac{\sigma}{1-\sigma}$ إلى يسارها، و عبارة عن نسبة خسارات الخطأ الإيجابي إلى الخطأ السلبي. وعلى سبيل المثال إذا كانت نسبة الخسارة Q = 4 فإن Z هي الدرجة التي تقع مساحة إلى يسارها 0.80 وقد استخدم نوفيك ولويس (Slwewis 1974 Wovick) خسارات التصنيف الخاطئ ونموذج بيتا ثنائي الحد في تطوير طريقة لإعداد معايير طول الاختبار ودرجات قطع التدريج الملاحظ.

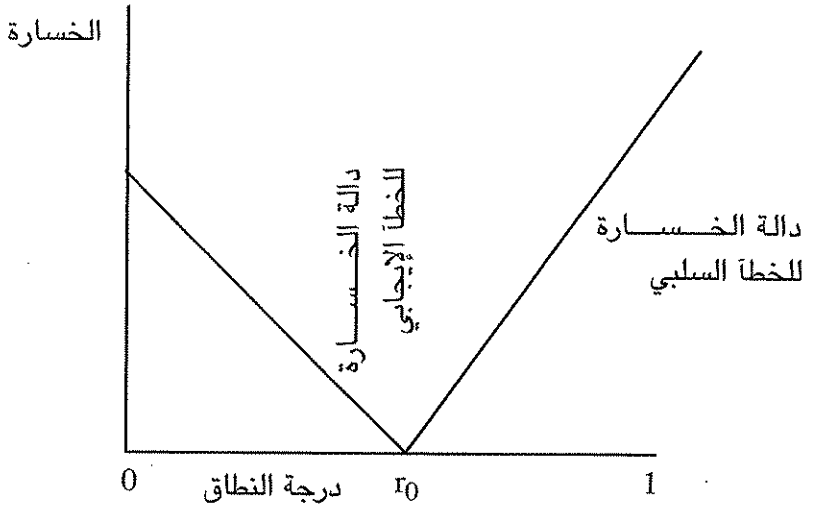
وفي التطبيقات السابقة لنظرية القرار تم معالجة التصنيف الخاطئ: الخطأ – السلبي جمعيها على أنها متساوية الجدية مثل التصنيف الخاطئ : الخطأ السلبي جميعها على أنها متساوية الجدية مثل التصنيف الخاطئ الخطأ الإيجابي.

تتضمن الخسارات من هذا النوع دالة فقدان العتبة (threshold loss Function) لذا فمن المنطقي الاستفسار عما إذا كان الخطأ ليس أكثر جدية في التصنيف الخاطئ لمفحوص مستوى قدرته الحقيقية بعيد جداً عن درجة القطع لتدريج النطاق أكثر مما هو لمفحوص آخر مستوى قدرته الحقيقية قريبة جداً من درجة القطع. للتوضيح، افترض أننا صنفنا الأطفال في برنامج تعليم القراءة، ونجح رالف في هذا الاختبار مع أن درجة النطاق له أقل بقليل من مستوى السيطرة، وكذلك نجحت سوزان في الاختبار ولكن درجة النطاق لها أقل بكثير من درجة القطع. وكان الفقدان المصاحب لقرار الخطأ الإيجابي لسوزان أكبر من الفقدان المصاحب لقرار الخطأ الإيجابي لرالف.

وإحدى دوال الخسارة التي تعبر عن هذه الأنواع من الخسارات دالة الخسارة الخطية. وهنا يكون الخسارة في خطأ الخطأ – السلبي يساوي $t_0 \geq Z > (t - t_0)a$ والخسارة في خطأ الخطأ السلبي يساوي $(t_0 - t)b$ ودوال الخسارة الخطية هذه موضحة في شكل (18–2) ويبين هذا التمثيل البياني الخسارة أن خطأ الخطأ– الإيجابي يوجد فقط لدرجات النطاق الأقل من t_0 (أي مفحوصين غير مسيطرين حقيقيين) . والخسارة لخطأ الخطأ – السلبي توجد فقط لدرجات نطاق أكبر أو تساوي t_0 (أي لمفحوصين مسيطرين حقيقيين) ولكلا النوعين من الأخطاء فإن درجة الخسارة تزداد كلما ازدادت المسافة بين t, t_0 وقد اشتق كل من فان دير لندن وميلنبرغ (Vanderlinden & Mellen bergh, 1977) درجة قطع التدريج الملاحظ لدوال الخسارة الخطية.

أن دوال فقدان العتبة والخسارة الخطية ليست الدوال الوحيدة الممكن اقتراحها. فأي دالة لـ t موجبة ولا تزيد عن صفر (أقل درجة نطاق ممكنة) إلى t_0 يمكن أن تخدم كدالة خسارة لقرارات الخطأ الإيجابي. وبصورة مشابهة الدالة الموجبة لـt والتي لا تقل عن t_0 إلى (وهي أكبر درجة نطاق ممكنة) يمكن أن تخدم كدالة خسارة لقرارات الخطأ – السلبي. وقد عدّ هونيه (Hunyh, 1980) طريقة تصغير القيمة الكبرى دالة خسارة غير خطية.

ومن المهم أن يتذكر الطبقتين أن تطبيق نظرية القرار المبينة هنا يفيد في إعداد درجة قطع تسمح لمستخدم الاختبار عمل استدلالات مناسبة عن حالة السيطرة الحقيقية للمفحوصين. ومع ذلك فمثل هذه الطريقة لا تقدم توجيهات للقضية الأولى لما يكون عليه مستوى الأداء لحالة السيطرة أو الطرائق التي يجب استخدامها لتحديد المعيار.



شكل (18-2) توضيح دالة الخسارة الخطية

كذلك لم نناقش قضية أي دالتي الاختيار سنستخدم، دالة فقدان العتبة أو ذاك الخسارة الخطية. كذلك لم نناقش اختيار دالة خسارة بعينها. على سبيل المثال وضع يتم الخسارات في مشكلة فقدان العتبة. بصورة أساسية ، يتطلب عمل مثل هذه الاختيارات لتقويم جدية الخسارات لكل درجة نطاق محتملة ومناقشة تفاصيل منهجيات مثل هذه الاختيارات خارجاً عن أهداف هذا الفصل. وعلى القارئ المهتم العودة إلى دراسة نوفيك ولندلي (Novick & Lindley , 1978) المراجع المذكورة فيها.

الخلاصة:

عندما تعتبر فقرات الاختبار ممثلة لنطاق الأداء يرغب الفاحص تأسيس درجة قطع ملاحظة أو معيارية والتي تسمح باستدلال من درجة المفحوص الملاحظة لحالة السيطرة الحقيقية في نطاق الأداء.

وهذا الحساب موضح في مواقف تستخدم فيها درجات الاختبار المكية لعلم قرارات حول شهادة أقل كفاءة. وفي هذه المواقف يجب أن يستخدم الفاحص أسلوب منطقي في تحديد درجة القطع والدفاع عنها. وهنالك ثلاثة طرائق سائدة تعتمد على تقديرات المحكمين على فقرات الاختبار منفردة هي طرائق نيدلسكي وانغوف وايبل. وبالإضافة إلى طريقة اقترحها جايجر والتي تجمع بين مظاهر طرائق أخرى عديدة بأسلوب التكرار أو التدوير . وتم مراجعة دراسات تجريبية قارنت بين المعايير الناتجة عن تطبيق طرائق متنوعة. وكان الاستنتاج العام منها

أن الطرائق المختلفة قد ينتج عنها معايير مختلفة وحتى مجموعات المحكمين المختلفة وباستخدام الطريقة نفسها يوصون بمعايير مختلفة. وفي ضوء هذا التباين نعرض بعض التوصيات العملية في إعداد المعايير:

1- السؤال عما إذا كان المعيار ضروري لموقف الاختبار.

2- حدد أكثر مهددات الصدق أهمية بالنسبة للاستنتاجات التي يمكن عملها باستخدام الاختبار.

3- اختر واعط تعليمات للمحكمين بطريقة تؤدي إلى خفض التحيز في أحكامهم.

4- استخدم طرائق متكررة في إعداد المعايير ومجموعات عديدة من المحكمين.

5- افحص المؤشرات التجريبية لكيفية أداء المجموعة النموذجية واستخدمها في تقييم توابع المعايير الموصى بها.

وهنالك فئتين من إعداد درجة قطع على التدريج الملاحظ، والتي تسمح لمستخدم الاختبار عمل استدلالات مناسبة لحالات السيطرة الحقيقية للمفحوصين إحداها تهتم بتقليل احتمالات التصنيف الخاطئ. وتكون هذه الطرائق مناسبة لحالات السيطرة الحقيقية للمفحوصين أحداها تهتم بتقليل احتمالات التصنيف الخاطئ وتكون هذه الطرائق مناسبة في حالة تساوي جدية الأخطاء : الخطأ الإيجابي والخطأ السلبي. وتهتم الفئة الثانية بتقليل الخسارات المتوقعة الناتجة عن التصنيف الخاطئ. وتستخدم هذه الطرائق دالة فقدان العتبة والتي تكون مناسبة في حالة كون أخطاء: الخطأ الإيجابي وأخطاء الخطأ السلبي متساوية في الجدية، ولكن أخطاء الخطأ الإيجابي والخطأ السلبي تكون مختلفة في جديتها. وتكون دوال الخسارة مناسبة في مواقف اخرى عرضت في هذا الفصل.

التمارين:

1) طبق اختبار مؤلف من (30) فقرة على (2000) طالب وكان المتوسط 18، و KR21= 0.6 فإذا كانت T0= 0.8، فما قيمة تقريب هونيه لـ X0 .

2) افترض للموقف الاختباري في التمرين الأجل أن نسبة النقد الخطأ الإيجابي إلى الخطأ السلبي = 2، فما قيمة تقريب هونيه لـX0.

3) باستخدام نتائج اختبار افتراضي مؤلف من (7) فقرات . حدد X0 إذا كان الفقد للخطأ الإيجابي (1) والفقد للخطأ السلبي (10) . حدد X0 إذا كان الفقد = 2 و 20

على التوالي. أيهما أكثر ضرورة في التطبيق وضع أخطاء الفقد للخطأ الإيجابي أم السلبي أم النسبة بينهما.

4) اقرا دراسة راولي الآتية Historical Antecedents of the Standard setting Debate : AnInside Account of the Minimal Beardedness Controversy Published in the Journal of Educational Measurement , 1982 , 19 , 87-96 .

وبعدها أجب عن الأسئلة الآتية:

أ- هل المتغير المهتم به كان من الممكن قياسه مفاهيمياً على مقياس متصل أم متقطع / نعم أم لا .

ب- زاوج بين طرائق إعداد المعايير المناقشة في هذا الفصل مع تلك المستخدمة في دراسة راولي وحدد على الأقل طريقتين من المعروضة في الفصل وفشل روالي في تناولها.

ج- هل من الممكن تطبيق طريقة تحكيم نيدلسكي للفقرة على صيغ الفقرات المقترحة في الدراسة.

د- حدد راولي واحد أو أكثر من النقد الرئيس أو القصور لكل طريقة من طرائق إعداد المعايير طريقة انغوف، وطريقة مجموعات المقارنة، ونظرية الاستجابة للفقرة. لكل طريقة صف نقد روالي وما مدى جديته فيما لو أسست درجة قطع لأقل كفاءة لشهادة تأهيل المعلمين المبتدئين.

هـ- من وجهة نظر نتائج راولي (وكما دونت في جدول 1 في هذه المقالة) هل من المنطق أن زيادة صدق المحتوى للفقرات (أو ملائمتها لنطاق مهم) يزيد من أقل درجة نجاح موصى بها لمجموعة الفقرات هذه. كيف تفسر هذه النتائج لدراسة راولي؟

الفصل التاسع عشر

19

المعايير والدرجات المعيارية

الفصل التاسع عشر

المعايير والدرجات المعيارية

ناقشنا في الفصل السابق الاختبارات محكية المرجع التي تهدف لان يكون للدرجات الخام معنى مفسر ومباشر. وفي العديد من الحالات من الصعب إجراء تفسيرات مفيدة أو استنتاجات من درجات المفحوصين الخام وحدها، وفي هذا الفصل سنعرض طرائق تدعم تفسير درجات الاختبار من خلال استخدام المعايير. وفي الحقيقة فإن الدرجات المعيارية جميعها تزودنا بمعلومات عن أداء المفحوصين مقارنة بتوزيع درجات عينة معيارية معينة أو مجموعة مرجعية، وبالتالي فإن معنى الدرجات يعتمد على 1) المدى الذي يهتم مطور الاختبار من خلاله مقارنة المفحوصين بالمجتمع المعياري 2) ومدى ملائمة العينة المعيارية لتمثيل المجتمع.

وفي البداية من المهم تمييز – ولبعض مستخدمي الاختبارات إن الأكثر ملائمة هو مقارنة أداء المفحوصين بأداء مجموعة الرفاق المعيارية (مثل المفحوصين الذين تقدموا للاختبار نفسه في الوقت نفسه) فاستخدام مجموعة الرفاق المعيارية يكون مناسباً لمستخدم الاختبار الذي يريد تعبئة شواغر يختارها من مجموعة متقدمين للوظيفة أو المعلم في الصف الذي يرغب في إعطاء درجات تعتمد على الأداء النسبي للطلبة في الصف الحالي وفي الفصل الدراسي الحالي. بالمقابل، يبدو أقل منطقية، وضع الطفل في برنامج علاجي معين أو عناية صحية متخصصة في مركز الولاية والتي تختلف اعتماداً على مجموعة معينة من المفحوصين تقدمت للاختبار في اليوم نفسه. لهذه الأهداف ما يساعدنا هو إمكانية مقارنة توزيع درجات الاختبار لعينة مفحوصين تمثل المجتمع المحدد بشكل جيد. ففي مثل هذه الحالات يجب وصف عينة المعايرة وصفاً دقيقاً بتفصيل كافٍ بالنسبة لخصائصها الديموغرافية (مثل الجنس، والسلالة أو الخلفية العرقية، والمجتمع أو المنطقة الجغرافية والحالة الاجتماعية الاقتصادية، والخلفية التربوية) وذلك لتسمح لمستخدمي الاختبار تقييم ما إذا كانت المقارنة ذات معنى مقارنة بأداء المفحوصين بمجموعاتهم المعيارية. وفي بعض المواقف يفضل مستخدم الاختبار تطوير معايير اعتماداً على أداء مجموعة معيارية محلية. وفي هذه الحالة يقع على عاتق مستخدم الاختبار مسؤولية بناء عينة معيارية مناسبة وإجراء تحويل الدرجات المناسب. وسنقدم في هذا الفصل بعض المفاهيم الأساسية المناسبة لتطوير معايير الاختبار أو تقييم ملائمة المعايير الموجودة مع

الأداة المنشودة، وبعدها سنراجع تحويلات الدرجات الشائعة المستخدمة في مقارنة درجة المفحوص الخام بتوزيع عينة المعايرة.

إجراء دراسة المعايرة:

بشكل عام، الخطوات الأساسية الموصى بها لإجراء دراسة المعايرة متشابهة عموماً بغض النظر عما إذا كانت المعايير محلية أو باستخدام موسع، وهذه الخطوات هي:

1. حدد المجتمع الذي تهتم بوضع معايير له (مثل طلبة المدارس العامة جميعها لإحدى الولايات أو المتقدمين للقبول في برنامج دراسي معين أو وظيفة معينة).
2. حدد الإحصائيات الأكثر أهمية التي سيتم حسابها لبيانات العينة (مثل المتوسط والانحراف المعياري والرتب المئينية).
3. قرر القدر المسموح به من خطأ المعاينة (الفجوة بين تقدير العينة ومعلم المجتمع) لواحد أو أكثر من الإحصائيات المحددة في الخطوة 2 (أهمها تحديد خطأ المعاينة للمتوسط).
4. اقترح طريقة لاختيار العينة من المجتمع (وسيتم وصف عدد من استراتيجيات المعاينة في الجزء التالي).
5. احسب اصغر حجم عينة مطلوب لجعل خطأ المعاينة ضمن الحدود المخصصة، وهنالك العديد من الصيغ يجب استخدامها والتي تعتمد على استراتيجية المعاينة المستخدمة.
6. اختر العينة واجمع البيانات. ووثق أسباب أي انتهاك قد يظهر، وعند ظهور انتهاك أساسي مثل تعذر اشتراك مدرسة كلياً بعد إن تم اختيارها في العينة. فمن الضروري استبدال هذه الوحدة باخرى يتم اختيارها بطريقة المعاينة نفسها.
7. احسب قيم احصائيات المجموعة وأخطائها المعيارية، وفي التطبيقات العملية فإن الخطأ المعياري للمتوسط هو المدون على الأغلب.
8. حدد أنواع الدرجات المعيارية التي تحتاجها، واجلب جداول تحويل الدرجات المعيارية اللازمة.
9. أعداد وثيقة مكتوبة لطريقة المعايرة والإرشادات لتفسير الدرجات المعيارية.

المعاينة الاحتمالية:

تدون البيانات المعيارية لدرجات الاختبار أحياناً بالاعتماد على بيانات عينات مناسبة (مجموعات سليمة من المفحوصين توافرت لطور الاختبار) ومعايرة الاختبار على عينة ما يزيد من احتمالية التحيز المنتظم الذي قد يؤثر على أداء المفحوصين، على سبيل المثال: المفحوصين

المتطوعين للاشتراك في معايرة استبانة مفهوم الذات قد يختلفون بانتظام في مفهوم الذات عن غير المتطوعين. إضافة إلى إن مثل هذه العينات لم يتم اختيارها لتمثل أي مجتمع، فإن استخدام احصائيات مثل الخطأ المعياري للمتوسط لتقدير درجة الخطأ المحتملة في التقديرات لا يكون لها معنى. بالمقابل عندما يستخدم مطور الاختبار عينة احتمالية لجميع بيانات المعايرة، فإن فرصة التحيز المنتظم الذي يؤثر في الأداء على الاختبار منخفض، إضافة إلى انه من الممكن تقدير كمية خطأ المعاينة المناسب الذي يؤثر على الاحصائيات المحسوبة المختلفة من هذه الدرجات. وببساطة تعرّف العينة الاحتمالية على أنها عينة يكون لكل فرد في المجتمع احتمالية معروفة لان يختار ضمن العينة، والصيغة الأبسط للمعاينة الاحتمالية (وهي التي تقارن بها الأنواع الأخرى عادة) هي المعاينة العشوائية البسيطة.

والمعاينة العشوائية البسيطة قد تشبه عملية تعيين رقم خاص بكل عضو في المجتمع، وكتابة كل رقم على قطعة ورق منفصلة ووضع قطع الورق جميعها في قبعة، ويسحب من القبعة عدد معين من الأوراق، وكل مفحوص يختار رقمه في هذه العملية يتم اختياره للعينة. وفي الواقع يتم اختيار نقطة بدء من جدول الأرقام العشوائية ويختار كل مفحوص يظهر رقمه حسب تتابع القائمة حتي يصل الباحث إلى العدد المطلوب من أفراد العينة.

وعندما تحسب قيمة متوسط الدرجات (أو أي إحصائي آخر) لعينة المعايرة، فإنه يعد تقديراً لمعلم المجتمع، ومثل هذا التقدير يكون خاضعاً لخطأ المعاينة، وعندما تتكرر سحب عينات عشوائية من مجتمع معين وسحب متوسط كل عينة، فمن الممكن وصف التوزيع التكراري لمتوسطات العينات حول قيمة واحدة لمتوسط المجتمع. وبالعودة إلى نظرية المعاينة فإن هذا التوزيع يقترب من المنحنى الطبيعي. والانحراف المعياري لهذا التوزيع لمتوسطات العينة يطلق عليه الخطأ المعياري للمتوسط. سنرمز له بالرمز σ_M وهذه الكمية تفيدنا عندما نريد تخصيص فترة حول المتوسط وبدرجة معينة من الثقة والمتوقع أن يكون متوسط المجتمع ضمنه. ومن الواضح انه كلما كانت الفترة اصغر، كلما كان الحكم على تمثيل عينة المعايرة للمجتمع أفضل.

ويمكن حساب قيمة σ_M من بيانات عينة واحدة من خلال الصيغة [1]

---

(1) تستخدم هذه الصيغة عندما يكون حجم العينة كبيراً حتى يمكن اعتباره يصل إلى اللانهاية، وعندما تتم المعاينة من مجتمع لانهائي فإن الصيغة الصحيحة هي :

$$\hat{\sigma}_M = \sqrt{\frac{\sigma_x^2}{N} \cdot \frac{(n-N)}{N}}$$

حيث N هو عدد الحالات في المجتمع. وفي معظم دراسات المعايرة فإن التصحيح لمجتمع لانهائي غير ضروري.

$$\sqrt{\frac{\hat{\sigma}_x^2}{n}} = \sigma_M \qquad (1\text{-}19)................$$

حيث ترمز $\hat{\sigma}_x^2$ إلى تباين درجات العينة [2]، وn هي حجم العينة. وواضح أن محددي الدقة لمتوسط العينة هو تباين المجموعة وحجم العينة. ويوضح جدول (19–1) كيفية انخفاض $\hat{\sigma}_M$ لمستوى ثابت من التباين بازدياد حجم العينة، وكيف يزداد لحجم عينة معين بازدياد التباين وطريقة أخرى لتقرير هذا هو أنه كلما كان التباين اكبر، كلما ازداد حجم

جدول (19-1) الأخطاء المعيارية للمتوسط لعينات مختلفة في تباينها وحجمها.

$\frac{\sigma_x^2}{n} = \sigma_M$	تباين العينة ($\hat{\sigma}_x^2$)	حجم العينة (n)
0.71	25	50
0.50		100
0.35		200
1.42	100	50
1.00		100
0.71		200
2.12	225	50
1.50		100
1.06		200

العينة اللازم لتحقيق مستوى معين من خطأ المعاينة, وعندما يستطيع مطور الاختبار تخصيص الكمية المسموح بها من الخطأ في المتوسط المحسوب لعينة المعايرة، فانه يمكن استخدام هذه المعلومات في صيغة الخطأ المعياري للمتوسط لحساب اصغر حجم يلزم من العينة. لذا ففي اختبار تحصيلي مؤلف من 80 فقرة فقد يرغب مطور الاختبار أن تكون درجة الثقة 95%، وهو إن أي تقدير للمتوسط يقع ضمن (∓ 1) نقطة لمتوسط الدرجة التي يمكن

(2) يجب ملاحظة انه عندما يحسب التباين من بيانات العينة، فان مجموع مربع انحرافات الدرجات يقسم على n –1 لا على n

الحصول عليها فيما لو اختبر المجتمع بأكمله. وهذا يعني أن نقطة التقريب تنتشر إلى σ_M^2، ويجب أن تكون أقل أو تساوي (1.00)، وبالتالي فإن $\sigma_M \geq 0.5$، ولتحديد اصغر حجم للعينة يلزم للحصول على درجة الدقة هذه فمن الضروري إن يكون لديك بعض الفكرة عن قيمة الانحراف المعياري لدرجات الاختبار (وهذه المعلومات يمكن الحصول عليها من خلال بيانات اختبار دليلي ابتدائي خلال جهود التطوير المبكرة أو من خلال عمل تخمين تربوي يعتمد على معرفة كيفية الأداء على اختبارات مشابهة من هذا النوع). افترض الآن أن مطور الاختبار لديه بعض الأسباب ليتوقع إن الانحراف المعياري على الاختبار التحصيلي هذا (المؤلف من 80 فقرة) سيكون 10 نقاط. وبتعويض هذه القيمة في المعادلة (19-1) وحلها لـ n، لدينا $n = \frac{100}{0.25} = 400$ لذا فإننا نحتاج عينة على الأقل مؤلفة من (400) فرد للحصول على درجة الدقة المرغوبة لمتوسط المجتمع من تقديرات العينة.

وفي بعض دراسات المعايرة من المستحيل سحب عينة عشوائية بسيطة، واختبار كل مفحوص مختار بهذه الطريقة. فعلى سبيل المثال ففي معايرة اختبار تحصيلي مقنن ليس من السهل أن تؤلف قائمة بأسماء طلبة الصف الثامن في الولاية وتختار بعضها عشوائياً لتطبيق الاختبار عليها، ويمكن تصور أن خطة المعاينة هذه قد تجعل الفاحص يزور عدد اكبر من المدن والمدارس والغرف الصفية في الولاية ويختار عشوائياً بعض الأفراد للاختبار. ولأهداف معايرة أخرى فان المعاينة العشوائية البسيطة قد لا تسمح للفاحص بضبط تركيب العينة لبعض المتغيرات المهمة والمناسبة. لذا فإن بدائل المعاينة العشوائية البسيطة قد تكون ملائمة أحيانا.

المعاينة المنتظمة:

تستخدم عندما يكون الأفراد مرتبين في قائمة، وهذا الترتيب غير مرتبط بالسمة المراد قياسها (مثل الأسماء في دليل الهاتف)، ويحدد الباحث نقطة بدء عشوائية من القائمة، وان كانت نسبة المعاينة المراد اختيارها $= ^1/_K$ من المجتمع، فإن كل فرد له الترتيب k في القائمة، ومن نقطة البدء المحددة يتم اختياره. وإذا وصل الباحث إلى نهاية القائمة قبل اختيار عدد أفراد العينة الكلي، ببساطة يذهب إلى بداية القائمة ويستمر في اختيار كل فرد له الترتيب K وتفيد المعاينة المنتظمة بشكل خاص عندما لا تتوافر قائمة بالأفراد كما هو الحال في مسح التسوق، إذ إن كل متسوق ترتيبه k. يتم اختياره في العينة بمجرد دخوله إلى المؤسسة أو يشتري منتج معين، ويمكن تطبيق هذه الطريقة في المعاينة في الحال لاختيار عينة المرضى أو

المسترشدين الذين يحتاجون برامج تشخيصية أو علاجية أو إعادة تأهيل، وان كان من الممكن افتراض إن قائمة الأفراد ترتيبهم عشوائي بالأساس على المتغير قيد الدراسة، فإن الخطأ المعياري لإحصائيات العينة يمكن تقديرها كما لو أنها عينة عشوائية بسيطة.

العينة العشوائية العنقودية:

تشبه المعاينة العشوائية البسيطة، ولكن يقرر الباحث مسبقاً إن العينة يجب إن تتضمن خصائص معينة لأنواع خاصة من المفحوصين. على سبيل المثال يرغب الباحث التأكد من إن العينة المختارة تتألف من 50% ذكور و 50% إناث أو 70% بيض و 30% سود. فان كانت n تمثل العينة الكلية فسيتم اختيار عينات الذكور والإناث في هذه الحالة باختيار عدد كلي يساوي (nx 0.50) ذكور (n x0.50) إناث وان كانت العينة 70% بيض و 30% سود و 50% ذكور و 50% إناث لكل مجموعة عرقية، وكان حجم العينة n فإن الباحث يختار

إحدى العينات العشوائية = n x 0.50 x 0.30 أو 0.15 n ذكور سود

واحدى العينات العشوائية = n x 0.50 x 0.30 أو 0.15 n إناث سود

وإحدى العينات العشوائية = n x 0.50 x .70 أو 0.35 n ذكور بيض

وإحدى العينات العشوائية = n x 0.50 x 0.70 أو 0.35 n إناث بيض.

وقد تكون المعاينة العشوائية العنقودية مفضلة على المعاينة العشوائية البسيطة لسببين، الأول لظهور المعاينة العنقودية عندما يكون المتغير مؤلف من طبقات وترتبط بالأداء على الاختبار.افترض على سبيل المثال لاختبار العلاقات المكانية أن هنالك اتجاه لان تكون درجات الذكور أعلى من الإناث وان النسب المؤلفة للمجتمع الأصلي تتألف من 70% ذكور و30% إناث في المجتمع الذي نهتم به. فإن سحبت عينة عشوائية تحوي 80% ذكور و 20% إناث فإن متوسط العينة سيكون اكبر من القيمة المقدرة لمتوسط المجتمع. وإن تم اختيار عينة عشوائية بسيطة تتضمن 50% ذكور و 50% إناث، فإن متوسط العينة سيكون أقل. كذلك فإن سحبت عينة عنقودية تتألف من 70% ذكور و 30% إناث فإن هذا سيزيد من أرجحية الحصول على متوسط عينة قريب من متوسط المجتمع. ويمكن لمطور الاختبار في العينة العنقودية ان ينتج معايير بخطأ معاينة أقل من تكلفة العينة العشوائية البسيطة نفسها وبحجوم متقاربة, وبشكل آخر فإن التقديرات من استراتيجيات المعاينة الاثنتين تتساوى فيها كمية الخطأ، ولكن المعاينة العنقودية ذات كلفة أقل، وبشكل عام يفضل استخدام المعاينة العنقودية عندما يكون تباين درجات المفحوصين ضمن الطبقة الواحدة اقل من تباين المجموعة

الكلي. كذلك فإن عينة المعاينة العنقودية تعكس التركيب الديموغرافي للمجتمع الذي يتسم بمصداقية اكبر من قبل مستخدمي الاختبارات فعلى سبيل المثال حتى في حالة عدم وجود اختلاف في أداء الطلبة الذكور والإناث في اختبار الاستعداد الأكاديمي للأطفال، فان مستخدم الاختبار يكون أكثر ارتياحاً عند مقارنته لدرجات المفحوصين على معيار يعتمد على عينة مؤلفة من 50% ذكور و 50% إناث أكثر مما لو كان معتمداً على استخدام مجموعة معيارية تتألف من 75% ذكور و 25% إناث، والتي لا تعكس الجنس في المجتمع العام.

وعندما تستخدم عينة عنقودية لمعايرة الاختبار فان متوسط العينة يطلق عليه أحيانا اسم المتوسط الموزون، والخطأ المعياري لهذا المتوسط يحسب بـ:

$$\sigma_{Mw} = \sqrt{3 W_h^2 \sigma M_h^2} \qquad (2\text{-}19)................$$

حيث يؤشر لكل طبقة بوزن W_h الذي يرمز إلى نسبة الحالات المتضمـنة في الطبقة [3] و σ_{Mh} يرمز إلى الخطأ المعياري لمتوسط الطبقة h.

وعندما يتجمع الأفراد معاً في تجمعات طبيعية أو وحدات تنظيمية (مثل الغرف الصفية، المصانع، المستشفيات، الأحياء السكنية) فمن الممكن بناء عينة باختيار عينة احتمالية من هذه الوحدات، وفي العينة العنقودية ذات المرحلة الواحدة، فحالما يتم اختيار الوحدة، فإن جميع الأفراد في تلك العينة يصبحوا من أفراد العينة، وسنحدد اعتبارنا هنا إلى الحالة التي يريد الفاحص بها معاينة التجمعات بطريقة يكون لكل تجمع احتمالية متساوية في اختياره ضمن العينة، ولسحب العينة فإن الفاحص قد يؤشر لكل تجمع أو طبقة بأرقام من (1 إلى k) ويتم اختيار التجمعات فيما بعد باستخدام جدول الأرقام العشوائية بالطريقة نفسها التي استخدمت المعاينة العشوائية البسيطة في اختيار الأفراد كوحدات منفصلة. وعادة يختلف حجم التجمع من طبقة لأخرى، وقد اقترح لورد **(Lord, 1959a)** أن الصيغة الملائمة لحساب الخطأ المعياري في هذا الموقف هو :

---

(3) عندما تكون المعاينة من مجتمع لا نهائي، فان الصيغة تعدل إلى:

$$\sigma_{Mw} = \sqrt{3 W_h^2 \sigma^2_{mh} (f_n - 1)}$$

حيث تعني f_h إلى كسر المعاينة، أو نسبة المفحوصين الذين اختيروا من الطبقة h.

$$\hat{\sigma}_{Mc} = \frac{\hat{\sigma}_{\mu}}{\sqrt{k}} \sqrt{2(\hat{\mu}_n / \hat{\sigma}_n) + 1} \qquad (3\text{-}19)\ldots\ldots\ldots\ldots\ldots\ldots$$

حيث ترمز k الى عدد المجموعات في العينة، و $\hat{\sigma}_M$ إلى الانحراف المعياري لمتوسطات المجموعة، و$\hat{\sigma}_M$ إلى الانحراف المعياري لحجم المجموعة، و $\hat{\mu}_M$ إلى متوسط حجم المجموعة، وفي حالة كون المجموعات جميعها متساوية في الحجم فإن $\hat{\sigma}_M$ تحسب بـ :

$$\hat{\sigma}_{Mc} = \frac{\sigma_{\mu}}{\sqrt{k}} \qquad (4\text{-}19)\ldots\ldots\ldots\ldots\ldots\ldots$$

ويكون المفحوصين ضمن المجموعة (مثل الصف) عادة أكثر تجانساً من المجتمع الكلي، لذا فإن متوسط عينة المفحوصين لعدد قليل من المجموعات يكون أبعد عن متوسط المجتمع من المتوسط الذي يعتمد على أعداد متساوية من المفحوصين الذين سحبوا بالمعاينة العشوائية البسيطة، وإذا أراد مطور الاختبار الحصول على إحصائيات معيارية لعينة عنقودية لها خطأ معاينة مشابه لما تم الحصول عليه من العينة العشوائية البسيطة لحجم عينة معين، فعليه إن يختار في حالة العينة العنقودية عدد اكبر من الأفراد. ووضح لورد (1959أ) إن المعاينة العنقودية في دراسة المعايرة لاختبار مقنن تتطلب اختبار مفحوصين يتراوح عددهم ما بين 12 إلى 30 ضعف للحصول على تقديرات معيارية قابلة للمقارنة (مقاربة) في الدقة للمفحوصين المختارين بناء على المعاينة العشوائية البسيطة. ويجب ملاحظة إن الفعالية الأكبر في المعاينة العنقودية يمكن تحقيقها بحيث تكون فرصة اختيار كل مجموعة يتناسب مع حجمها، وقد وصف جايجر (Jaeger,1984) طرائق المعاينة وزودنا بصيغ للحساب في أسلوب المعاينة العنقودية هذه.

ولان الأخطاء المعيارية لإحصائيات التقنين تكون عموماً اقل بزيادة عدد الطبقات المستخدمة، يجب إن يخطط الفاحص لجمع بيانات من اكبر عدد ممكن من الطبقات.وإحدى الطرق لاحتواء التكلفة الكلية للدراسة (وبالتالي إمكانية الحصول على مدى استجابات أفضل من الوحدات المشاركة) من خلال المعاينة متعددة المراحل، والتي توضح بسهولة كبيرة من خلال المعاينة ثنائية المرحلة، افترض إن مطور الاختبار كان مهتماً في تطوير معايير لمقاطعة محلية لاختبار الاستعداد الأكاديمي المطبق فردياً على الطلبة في مستوى دراسي معين، ولكنه لا يستطيع اختبار طلبة المقاطعة جميعها، هنا يمكن استخدام تصميم المعاينة ثنائية المرحلة وتكون كما يأتي:

1. اختيار عينة عشوائية من المدارس.

2. ضمن كل مدرسة اختيار نسبة ثابتة من المفحوصين (مثلاً 5%)، وهذه الطريقة تمنح كل مدرسة فرصة متساوية للاختيار ولكن التأكيد على إن المدارس الكبيرة تمثل ثقل اكبر في العينة من المدارس الصغيرة، ويتأثر الخطأ المعياري للمتوسط للعينة ثنائية المرحلة بخطأ التباين الذي يعود إلى معاينة المدارس وخطأ التباين الذي يعود إلى معاينة الأفراد ضمن المدارس، وخطأ التباين الذي يعود إلى معاينة الأفراد ضمن المدارس، يمكن حسابه بـ $\hat{\sigma}^2_{Mc}$ (انظر المعادلة 19-3)، وقد حسب كل من انغوف (Angoff,1971) ولورد (lord, 1959a) خطأ التباين الذي يعود لمعاينة الأفراد ضمن المجموعة باستخدام صيغ مكافئة لـ :

$$\frac{\hat{\sigma}_x^2 \ (P-1)}{\hat{\mu}_n \, pk} = \sigma^2_{Mw}$$

وفي مثالنا، أ ترمز إلى نسبة الطلبة الذين تم اختيارهم من كل مدرسة، $\frac{k}{\hat{\sigma}_2}$ هي عدد المدارس، $\hat{\mu}_n$ هي متوسط حجم المدرسة، $\hat{\sigma}_x{}^2$ هي تقدير لمتوسط التباين ضمن المدرسة للمجتمع، لذا فان الخطأ المعياري للمتوسط لهذه العينة ثنائية المرحلة هو:

$$\sigma^2_{mE} + \sigma^2\mu_{c} = \hat{\sigma}_M \qquad (5\text{-}19)................$$

ما سبق ليس تصميم المعاينة الوحيد وصيغ التقدير الخاصة به لاختيار العينة ثنائية المرحلة للمجموعات والأفراد، أسلوب اخر ممكن هو اختيار عينة المجموعات من المدارس بطريقة معينة يكون فيها فرصة اختيار المدرسة يتناسب مع حجمها وبعدد متساوي من التلاميذ الذين سيختبرون من كل مدرسة، وقد وصف انغوف (1977) طريقة عملية في اختيار مثل هذه العينة ثنائية المرحلة، كذلك فإن كل من انغوف ولورد قدما صيغة لحساب الخطأ المعياري لمتوسط مثل هذه العينات. والنقطة الرئيسية هنا انه يجب على القارئ إن يميز انه في كل مرحلة من المراحل المتتالية في عملية المعاينة يضاف عنصر خطأ تباين يؤثر على متوسط الخطأ المعياري. وعندما تكون المجموعات المستخدمة سليمة في تركيب عينة المعايير فإنه لا يجب معالجتهم على أنهم عينة عشوائية من الأفراد سحبت من المجتمع إضافة إلى أن الستراتيجيات المختلفة لمعاينة المجموعات والأفراد تتطلب صيغاً مختلفة لتقدير الخطأ ويمكن الإفادة من مرجع كيش(kish,1965) في بناء عينات متعددة المراحل.

وصف دراسة المعايرة في دليل الاختبار:-

على مطور الاختبار مسؤوليات عديدة في وصف نتائج دراسة المعايرة. الأول يجب وصف المجتمع العام للمفحوصين المستهدفين في الاختبار، وبعدها وصف وتوثيق للطريقة التي

اختيرت بوساطتها عينة المعايرة وهذا يتضمن بالتأكيد وصف تفصيلي لخطة المعاينة (مثل نوع المعاينة المستخدم، وتحديد الطبقة في المعاينة الطبقية، ونسبة المعاينة لكل طبقة، وعدد المجموعات المختارة). وكاقتراح عام يجب أن يكون وصف خطة المعاينة شاملاً لدرجة كافية بحيث يمكن لمستخدم الاختبار أن يحسب الخطأ المعياري لإحصائيات (غير تلك المدونة في الدليل).

وفي وصف طريقة المعاينة فمن المهم الإشارة إلى مدى الرفض أو عدم الاستجابة عبر وحدات المعاينة الأصلية، ولتفسير التأثيرات المحتملة، ومثل هذا الرفض لا يمكن تعميم نتائجه من العينة إلى المجتمع. فعلى سبيل المثال مع أن دليل اختبارات الميتروبوليتان التحصيلية (prescotietal, 1978) يشير إلى أن أكثر من 550.000 طالب اشتركوا بشكل ما في المعايرة، وأشار باغلين (Bagling,1981) إلى إن ما مجموعه 88 مدرسة من المقاطعات وافقت بالفعل على الاشتراك بعملية المعايرة، إلا إن هذه المقاطعات تتضمن 32% فقط من المقاطعات التي تم دعوتها للاشتراك في دراسة المعايرة الوطنية: (مدى مشابه أو أقل من المشتركين تم تدوينه في معايرة اختبارات تحصيلية أساسية أخرى خلال الفترة نفسها). إضافة إلى إن الرغبة للاشتراك في دراسة معايرة لناشر معين يبدو أنها تتأثر بسلسلة الاختبارات التحصيلية التي تستخدم في الوقت نفسه في المقاطعة. لذا تساءل باغلين عما إذا كان من المنطق وصف مثل هذه المعايير على أنها في الحقيقة تمثل معايرة وطنية، مقترحاً على ناشري الاختبارات بان يكونوا أكثر شمولاً في وصفهم لماهية بناء أو تركيب عينات المعايرة.

معلومات أخرى حاسمة يجب تدوينها تتضمن تاريخ دارسة المعايرة مع الوصف لتركيب العينة المعيارية من خلال الجنس والخلفية العرقية أو السلالة، والحالة الاجتماعية الاقتصادية، والموقع الجغرافي، وأنواع المجتمعات الممثلة – حينما يكون مناسباً – وأي ظروف خاصة طبق الاختبار تحت شروطها فيجب ذكره. فعلى سبيل المثال، إذا كانت عينة المعايرة لامتحان الكفاءة التخصصية تتألف من المتقدمين لامتحان الكفاءة والذين لزم تطبيق الاختبار عليهم على أساس تجريبي فقط (لذا لا تؤخذ درجاتهم بعين الاعتبار في قرار الكفاءة، فمستوى دافعيتهم وبالتالي اداءهم الاختباري قد يكون أقل من الذين سيختبرون لاحقاً والذين يجب إن يجتازوا الامتحان ليحصلوا على الكفاءة.

ويجب أن يرافق إحصائيات المجموعة المدونة لوصف أداء العينة المعيارية على الاختبار (مثل المتوسط والانحراف المعياري) معلومات تشير إلى دقة تقديرات العينة. فعلى الأقل يجب تدوين الخطأ المعياري للوسط يصاحبه فترات درجة ثقة معينة تتضمن متوسط المجتمع لمستويات مختلفة من الدقة. وقدم انغوف طريقة تجريبية لتقييم ما إذا كان الخطأ المعياري

للمتوسط ضمن حدود منطقية باقتراحه أن الخطأ عند المتوسط يعود إلى تجميع خطأ المعاينة، وان الخطأ المعياري للاختبار يجب أن لا يتجاوز الخطأ المعياري للقياس بأكثر من 1%.

أخيرا، يجب إن يتضمن دليل الاختبار تفسيرات واضحة للمعاني، والتفسيرات المناسبة لكل نوع من تحويلات الدرجات المعيارية. وهنا يفضل الإشارة إلى كم التغير المتوقع لكل نوع من الدرجات بسبب أخطاء القياس وللتحولات مثل الرتب المئينية يجب توضيحها عند نقاط مختلفة من التوزيع.

أنواع الدرجات المعيارية

تذكر إن معظم حالات دراسة المعايرة تجري بهدف بناء جداول تحويلات بحيث إن قيمة درجة خام معينة يمكن تفسيرها من خلال موقعها النسبي وتكراراتها ضمن توزيع الدرجات الكلي وسنصف في الأجزاء التالية الأنواع الشائعة من الدرجات المعيارية مع تبيان ما يجب إن يؤخذ بالحسبان عند تفسيرها.

الرتب المئينية

مجمل القول إن الرتب المئينية المناظرة لدرجة خام معينة تفسر على أنها نسبة المفحوصين في مجموعة المعايرة الذين تقل درجاتهم عن هذه الدرجة، وحسابياً تحدد الرتب المئينية على أنها :

$$\% \quad 100 \text{ x } \frac{(fi)\ 0.5}{N} \quad + C_{fi} = P \qquad (6\text{-}19)................$$

حيث ترمز C_{fi} إلى التكرار التراكمي للدرجات جميعها التي تقع دون الدرجة، و f_i إلى تكرار الدرجة المراد حساب رتبتها المئينية، وN إلى عدد أفراد العينة، وخطوات حساب الرتب المئينية لتوزيع الدرجات الخام هي:

1/ تكوين جدول تكراري للدرجات الخام كما في جدول (19-2).

2/ ولدرجة خام معينة، حدد التكرار التراكمي للدرجات التي تقع دون الدرجة المعينة (فإن أردنا حساب الرتبة المئينية المناظرة للدرجة 17 من بيانات جدول (19-2) فعلينا تحديد التكرار التراكمي للدرجة 16.

جدول (19-2): التوزيع التكراري للدرجات الخام،

والرتب المئينية والدرجات الخطية والدرجات الزائية المعيارية القياسية

الدرجة الخام	f	cf	الرتبة المئينية (a)	الدرجة الزائية الخطية	الدرجة الزائية المعيارية
11	2	2	1	2.53-	2.33-
12	1	3	2	-2.17	2.05-
13	6	9	4	-1.8	1.75-
14	5	14	8	-1.44	1.40-
15	12	26	13	-1.07	1.13-
16	17	43	23	-0.71	0.74-
17	21	64	36	-0.34	0.36-
18	28	92	52	0.02	0.05
19	19	111	67	0.39	0.47
20	15	126	79	0.75	0.81
21	10	136	87	1.12	1.13
22	5	141	92	1.48	1.40
23	3	144	95	1.85	1.64
24	4	148	97	2.21	1.88
25	2	150	99	2.58	2.33

(a) مقربة لأقرب منزلتين عشريتين. μ_i: 17.94 ، $\hat{\sigma}_X$ =2.74

3. أضف نصف تكرار الدرجة المراد حساب رتبتها المئينية في الخطوة 2

4 . اقسم المجموع على N، عدد المفحوصين في عينة المعايرة واضربها بمئة.

والرتبة المئينية المناظرة للدرجة الخام 17 في جدول (19-2)، تحسب كما يأتي:

$$PR17 = \frac{(21)\ 0.5 + 43}{150} \times 100\% = 36$$

لغاية الآن ليس واضحاً لماذا نأخذ نصف عدد المفحوصين الذين حصلوا على الدرجة الخام نفسها في حساب الرتبة المئينية، كذلك فان الدرجات الخام عبارة عن قيم متقطعة،

وتقليدياً تعد السمة المقيسة في الاختبار متصلة القيم (تقع على تدريج متصل)، لذا فان كل قيمة للدرجات الخام تناظر فترة للدرجة على متصل القدرة أو السمة المقيسة، وتتركز نقطة الاهتمام على وسط هذه الفترة. ونظرياً فإن نصف المفحوصين الذين حصلوا على درجة خام معينة تعد درجاتهم على السمة اقل من نقطة وسط الفترة. والنصف الآخر تعد درجاتهم أعلى من نقطة وسط الفترة. لذا فإنه عند حساب الرتبة المئينية فإننا نحسب (نعد) عدد المفحوصين الذين هم دون نقطة وسط الفترة (أي نصف عدد المفحوصين الذين حصلوا على الدرجة الخام نفسها).

ويوصى باستخدام الرتب المئينية عند إيصال نتائج الاختبارات معيارية المرجع إلى التلاميذ والوالدين والمرشدين. فإذا افترضنا أن مجموعة المفحوصين هي مجموعة معيارية، وتمت مراقبة تطبيق الموقف الاختباري ووقت الاختبار بعناية، فإنه يمكن توضيح التفسير الملائم بهذه الصيغة، درجة جين في الاختبار أعلى من 75% من المجموعة المعيارية، ومع انه يوجد عدة محاذير يجب إن تؤخذ بعين الاعتبار عند تفسير الرتب المئينية. فمعظم أخطاء التفسير تظهر عندما يفشل مستخدم الاختبار تمييز إن تدريج الرتب المئينية هو تحويل غير خطي لتدريج الدرجات الخام. وببساطة فهذا يعني انه على المناطق المختلفة من تدريج الدرجة الخام كسب نقطة واحدة يناظرها كسب مختلف القيمة على تدريج الرتب المئينية.

للتوضيح، دعنا نتفحص جدول (19-3) الذي يتضمن أمثلة لدرجات خام وقيم الرتب المئينية المناظرة وذلك لاختبار تحصيلي في الرياضيات للمرحلة الأساسية، الخطأ المعياري له =3 نقاط تقريباً. وبالتالي فإن العمود الرابع من الجدول يتضمن فترات الدرجة الخام المناظرة لـ: ($1\sigma_E$+ x)، لاحظ انه عندما يتساوى عرض الخلايا لقيم الدرجات الخام جميعها فإن فترات الرتب المئنية المناظرة تختلف اختلافاً ملحوظاً بالاعتماد على موقع الدرجات الخام في التوزيع. ويظهر هذا لان التكرار النسبي الأكبر للدرجات في وسط التوزيع، والانخفاض النسبي للتكرار بالاتجاه نحو ذيل التوزيع. لذا فإن مستخدم الاختبار يجب إن يكون واعياً لـ:

1/ إن درجات الرتب المئينية أقل استقراراً (ثباتاً) للدرجات التي تقع في مركز التوزيع أكثر مما هو عند الإطراف. فعلى سبيل المثال في جدول (19-3) نرى إن درجة مفحوص متوسط القدرة قد تتغير بـ 12 رتبة مئينية وتبقى غير متجاوزة للخطأ المعياري للقياس، في حين إن الزيادة نفسها في الدرجة الخام لمفحوص عند مدى القدرة العلوي ينتج عنه تغير قدره خمس رتب مئينية، ومع ذلك فإن نطاق المئين (هو نطاق الرتبة المئينية المناظرة للنقاط $1\sigma_E$ + x) لا يزال مفيد في تفسير الدرجة. ومثل هذا النطاق للرتبة المئينية مبين في أقصى عمود إلى اليمين في جدول (19-3).

2/ إن الكسب أو الخسارة في الرتب المئينية للمفحوصين المنفردين لا يمكن مقارنتها

معنوياً لمفحوصين عند نقاط مختلفة من التوزيع. فعلى سبيل المثال، افترض ان التلاميذ في برنامج تجريبي اكتسبوا جميعاً (3) درجات خام في الاختبار البعدي مقارنة بدرجات الاختبار القبلي. هنا مقارنة الكسب من خلال الرتب المئنية يؤدي إلى استنتاج غير ملائم وهو إن فئة المفحوصين ذوو القدرة الأعلى كان اثر المعالجة عليهم أقل مما هو للمفحوصين في وسط التوزيع إذ أنهم حصلوا على اكبر قدر من الكسب

3/ لا يمكن تفسير الحسابات الرياضية والإحصائية لدرجات الرتب المئينية بمعنى أو هدف في بعض المواقف.

جدول (19-3): توضيح للدرجات الخام والرتب المئينية

ونطاق الرتب المئينية لاختبارتحصيلي في الرياضيات في المرحلة الأساسية

منطقة التوزيع	الدرجة الخام	الرتب المئينية	$1\ \sigma_E \pm X$	نطاق الرتبة المئينية $1\sigma_E \pm X$
الأعلى	41	94	38-44	84-99
الوسط	28	38	25-31	27-50
الأدنى	20	14	17-23	8-22

(أ) : $\hat{\sigma}_E$ تقريباً = 3 نقاط.

افترض على سبيل المثال أن الباحث أراد مقارنة متوسطات مجموعتين. وقد يختلف الاستنتاج حول أداء المجموعتين فيما لو استخدم متوسطات الرتب المئينية بدلاً من استخدام الدرجات الخام، ويمكن رؤية هذا من خلال مثال بسيط لبيانات جدول (19-2) افترض أن المجموعة أ تألفت من مفحوصين اثنين درجاتهم الخام (12 , 20)، والمجموعة ب تألفت من مفحوصين اثنين درجاتهم الخام (15 و 17) وكلا المجموعتين متوسطهما = 16 نقطة، ولكن متوسطات درجات الرتب المئينية المناظرة مختلفة اختلافاً ملحوظاً (40.5 للمجموعة أ، و24.5 للمجموعة ب)، لذا فإن الاعتماد على متوسط درجات الرتب المئينية يعد غير مناسب لمقارنة المجموعتين.

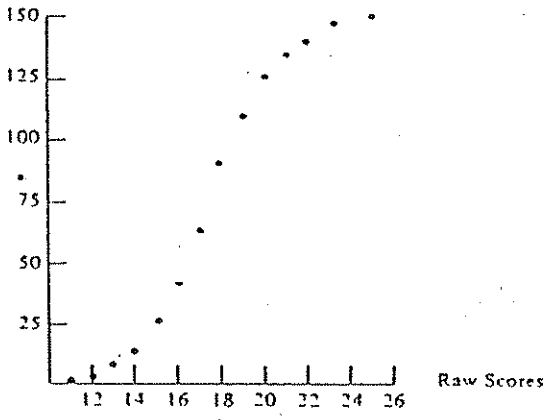
الدرجات الزائية المعيارية :

مع إن معظم توزيعات الدرجات الاختبارية ليست طبيعية بشكل مطلق (LORD,1955) إلا

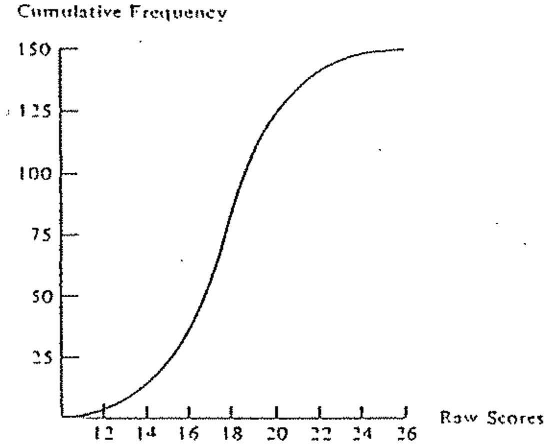
إن هنالك بعض الحسنات لتفسير درجات الاختبار المرتبطة بإنشاء درجات معيارية اعتماداً على الاعتدالية (التسوية الاعتدالية) لتوزيع الدرجات الأصلي. إن ملائمة أو مناسبة تدريج Z المعياري انه وبغض النظر عن عينة الاختبار أو المفحوصين، فإنه يوجد لكل نقطة على التدريج نسبة مئوية ثابتة للحالات التي تقع أعلى أو أدنى من تلك النقطة، ويمكن الحصول على قيم النسب المئوية من جداول درجات Z المعيارية. ويتم الحصول على درجات Z المعيارية بإجراء تحويل غير خطي للدرجات الخام الأصلية، وبخلاف تحويل درجات Z الخطية (المذكورة في الفصل2) التي تحتفظ بتوزيع الدرجات الخام، فإن درجات Z المعيارية لها توزيع يقارب المنحنى الاعتدالي. لذلك فإن درجات Z المعيارية ودرجات Z الخطية للدرجات الخام نفسها تختلف إلى الحد الذي تبتعد فيه الدرجات الخام عن الاعتدالية..

ولتحويل الدرجات الخام إلى درجات معيارية بأسلوب بسيط نحدد الرتبة المئينية المناظرة لكل درجة خام بالطريقة التي ذكرت في الجزء السابق، ومن ثم النظر إلى درجة Z الرتبة المئينية وذلك باستخدام جدول درجات Z المعيارية القياسية، وكما تم في جدول (19-2) فإن درجات Z المحددة لقيم درجات خام متنوعة تتغير من عينة لأخرى اعتماداً على تباين المعاينة في تكرارات درجة خام معينة. لذا فإنه يستخدم في دراسات المعايرة واسعة النطاق أسلوب أكثر تعقيد في تحديد الرتب المئينية ودرجات Z المعيارية، وهذا الأسلوب يستخدم عملية مطابقة للمنحنى تسمى تمليس (SMOOTHING) والتي تتغير بشكل ملحوظ. لذا فعلى القارئ إن يميز إن العرض هنا هو فقط لتحديد المفهوم العام.

افترض انه تم تمثيل الدرجات الخام على المحور الأفقي في التمثيل على المستوى البياني، وتمثيل الرتب المئينية المرتبطة بالدرجات الخام على المحور العمودي (كما في شكل9-1أ) ويكون المنحنى الناتج عبارة عن رسم لنقاط متقطعة، ومع ذلك فمن الممكن تثبيت منحنى بشكل S لمجموعة النقاط هذه (انظر شكل 19-1 ب)، فإن كان توزيع الدرجات الخام طبيعي تماماً فإن المنحنى الناتج يكون طبيعي (نوقش في الفصل 15) وهنالك أسلوبين عامين يمكن استخدامهما لتمثيل منحنى أملس لمنحنى النقاط. احدهما هو رسم منحنى باليد باستخدام منحنى مثل المبين في شكل 19 –1 أ، ومع ذلك فقد اقترح انغوف (Angoff, 1971) استخدام ورقة الاحتمالات الطبيعية (normal probability paper) وهي ورقة بيانية تبنى بحيث ترسم النقاط من التوزيع الطبيعي التي تقع على خط مستقيم واحد. والأسلوب الثاني يستخدم المطابقة مع دالة رياضية للتوزيع التكراري، وبناء توزيع تكراري تراكمي من الدالة الرياضية.



(أ) تمثيل بياني للدرجات الخام والتكرار الراكمي



(ب) منحنى أملس بالاعتماد على التمثيل البياني

شكل(19-1) الأزواج المرتبة ومنحنى للعلاقة بين الدرجات الخام والتكرارات التراكمية لبيانات جدول 19-1

وقد أوضح كل من كيتس ولورد **(Keats & Lord,1962)** ولورد ونوفيك **(Lord & Novick ,1968)** تطبيق دالة فوق هندسية سالبة لهذا الهدف.

وحالما يتم رسم المنحنى الأملس الذي يبين العلاقة بين القيم على تدريج الدرجة الخام والقيم على تدريج الرتبة المئينية، فإن الرتبة المئينية المناظرة لدرجة معينة تؤخذ عن المحور الرأسي وتعتمد بدلاً عن الرتبة المئينية الأصلية المحسوبة من تكرارات البيانات الملاحظة،

وبالتالي فإن درجة Z المعيارية القياسية تحدد أيضا بإيجاد درجة Z من جدول المنحنى الطبيعي التي تناظر درجة الرتبة المئينية الملساء. ويجب ملاحظة إن التمليس (SMOOTH-ING) عملية عامة في بناء التدريج وليست محددة بالمنحنيات التي تبين العلاقة بين الدرجات الخام والرتب المئينية ويمكن تطبيقها على تحويلات التدريج المختلفة

إن قرار معايرة الدرجات القياسية في تطوير المعايير ليست واضحة بشكل قطعي. وتاريخياً فإن هذا التطبيق أُثري بشكل مناسب خاصة في تطوير مقاييس الاستعداد والتحصيل المنتشرة تجارياً، ولمثل هذه الاختبارات فإن هذه النقطة خاضعة للمناقشة لأنه في العديد من الأمثلة فان التطبيقات في كتابة الفقرات وفي اختيارها تستخدم لإنتاج اختبارات شريطة إن يكون توزيع مجتمع المفحوصين طبيعياً، وحجوم العينات كبيراً لدرجة كافية ليمثل توزيع المجتمع بدقة. وتصبح عملية المعايرة أكثر أهمية وذلك عندما يبتعد توزيع الدرجات بصورة أساسية عن التوزيع الاعتدالي. وهذا شائع في أبحاث التدريج الصغيرة أو دراسات المعايرة المحلية حينما يكون حجم عينات المعايرة صغير بشكل ملحوظ. وفي مثل هذه الحالات فإن توزيع درجات Z المعيارية قد يختلف اختلافاً واضحاً عن توزيع الدرجات الخام الأصلية. وهنا يبرز سؤال عما إذا كانت الدرجات الخام أو الدرجات المعيارية هي التي تمثل توزيع الدرجات بشكل أكثر دقة (أكثر ثقة) مما لو اختبر المجتمع الأصلي برمته. من جهة أخرى يبدو إن بعض منظري القياس يؤيدون أو يدافعون عن معايرة الدرجات وأن لها مستوياتها، وذلك في حالة اعتقاد الباحث أن السمة النفسية المستهدفة تتوزع طبيعياً في المجتمع (Magnesson,1967). وادعى جوليكسن(Gulliksen,1950) انه من المنطق الاعتقاد بان القدرة المقيسة بالاختبار تتوزع طبيعياً ولكن الانحرافات في الاختبار سببت انحراف الدرجات الملاحظة عن الاعتدالية. ومن جهة أخرى حذر انغوف (1971) من التحويل إلى التوزيع الاعتدالي الذي قد يكون غير مرغوب، خاصة إذا كان شكل التوزيع غير الاعتدالي دالة لعملية اختيار المفحوصين، أو أي توزيع غير مألوف للقدرة في المجموعة المعينة التي تم اختبارها. لذا فإنه لأمر متشدد نوعاً ما القول بأن معايرة الدرجات لها مستوياتها عندما لا يكون انحراف الدرجات متطرف جداً عن الطبيعي لنبدأ به ومن الصعب تخيل موقف يكون توزيع درجات العينة بعيد كثيراً عن التوزيع الطبيعي، وبالتالي فإن الفاحص يقتنع بأن العينة هي تمثيل مناسب لمجتمع ما يكون توزيع السمة فيه طبيعياً.

تحويلات الدرجات الزائية:

في كلا الحالتين الخطية، وغير الخطية، فإن الدرجات الزائية لها مساوئ من حيث استخدام قيم عشرية (كسور) أو سالبة أو كلاهما والتي يصعب التعامل معها حسابياً وصعوبة تفسيرها

لمستخدمي الاختبارات (افترض لوهلة مهمة محاولة التفسير لمفحوص تقدم لاختبار من 300 فقرة وأجاب عن 100 منها إجابة صحيحة وحصل على الدرجة Z = 1.5-). يبدو انه أكثر ملائمة إجراء تحويل خطي للدرجات الزائية لتحويلها إلى قيم أسهل تسجيلاً وتفسيراً. ومن المهم أن نتذكر (وكما لاحظنا في فصل 2) إن التحويلات الخطية لا تغير شكل توزيع الدرجات الزائية. والصيغة العامة لمثل هذا التحويل هو.

$$(Z)\,k + m = y \qquad \text{..............} (7\text{-}19)$$

حيث ترمز Y إلى الدرجة المحولة والقيم m و k قيم ثابتة تختار اصطلاحاً من قبل مطور الاختبار والقيمة m تمثل متوسط التوزيع الجديد بعد التحويل، والقيمة k إلى الانحراف المعياري الجديد .

ويمثل الشكل (19-2) عدة تحويلات شائعة لدرجات Z تحت المنحنى الطبيعي. وأحد من التحويلات هو الدرجات التائية التي طورها مكول (Mc call، 1939) والذي اقترحه لتدوين اداء الاطفال على اختبار القدرات العقلية والصيغة العامة للدرجات التائية هو:

$$50 + 10\,(Z) = T \qquad \text{...............} (8\text{-}19)$$

والتدريج التائي يكون يكون وسط المجموعة المعيارية = 50، وانحرافها المعياري = 10 وتحويل اخر للدرجات الزائية وينصح به من قبل خدمات الاختبارات التربوية (Educ Teating Service) لتدوين درجات القبول في الكلية العام يأخذ الصيغة :

$$500 + 100\,(Z) = y \qquad \text{...............} (8\text{-}19)$$

لذا فان متوسط توزيع الدرجات المحوّلة =500، وانحرافه المعياري = 100 نقطة ومن اكثر التحويلات شهرة للدرجات الزائية جميعها هو انحرافات درجات 1.Q والتي لها الصيغة العامة:

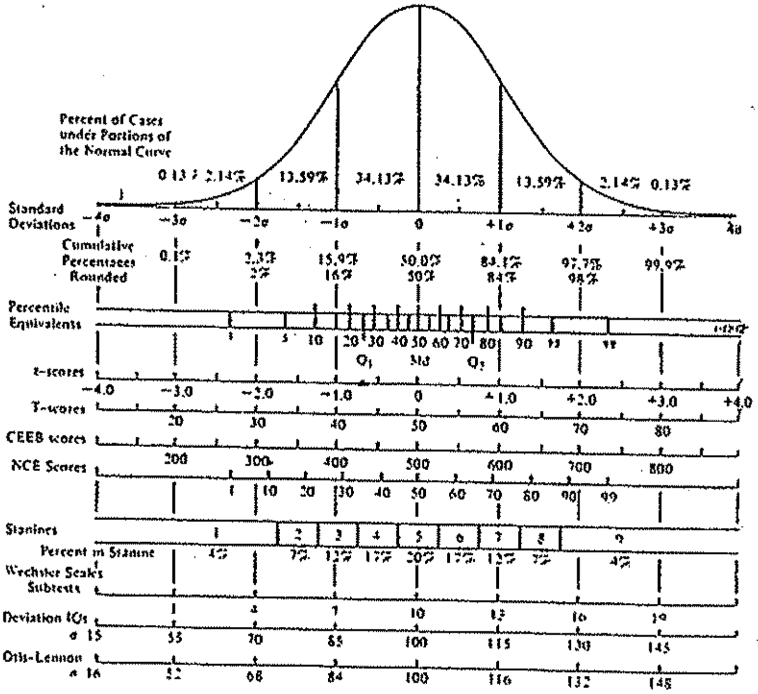
$$100 + 15\,(Z) = DIQ \qquad \text{..............} (10\text{-}19)$$

والتي استخدمها وكسلر (1939) في تفسير الدرجات على مقياس ذكاء البالغين. وتدريج الدرجات المشتق هذا صمم ليكون متوسطه 100 نقطة وانحرافه المعياري 15 نقطة. وتستخدم اختبارات الذكاء الاخرى مثل مقياس ستانفورد – بينيه انحراف معياري 16 نقطة. وللدقة التاريخية يجب ملاحظة ان المفهوم الاصلي لحاصل الذكاء quotient intelligence . اقترحه

ويليهم (Wilhelm,1982) لاستخدامه في مقياس بينيه، وهو ليس درجات IQ التي قدمها كابلان وساكوزو (Kaplans & saccuzzo , 1982) الذين لهم الفضل لهذا التطور التاريخي لإختبارات الذكاء.

اخيراً، تدون بعض ادلة الاختبارات التحصيلية المقننة جداول لتحويل الدرجات الخام او الرتب المئينية الى مكافئات على المنحنى الطبيعي (NCE)، والتي تكون مرتبطة بشكل خاص بنموذج التقويم معياري المرجع (Tallmadge & wood ,1976) وهذا التدريج له متوسط (50) نقطة وانحراف معياري (21.06) نقطة وتستخدم الجداول المعيارية في تحويل الرتب الميئنية الى تدريج مكافئات المنحنى الطبيعي (NCE) والتي يتم الحصول عليها بتحديد الدرجات الزائية المعيارية القياسية المرتبطة بالرتب المئنية واجراء تحويل باستخدام الصيغة :

$$50 + 21.06\ (Z) = NCE \qquad (11\text{-}19)................$$



شكل (2-19) : المنحنى الطبيعي، المئينات، ودرجات معيارية منتقاة اعيد طباعتها من مفكرة خدمات الاختبار 148 لرابطة السيكولوجيين.

التدريج التساعي:

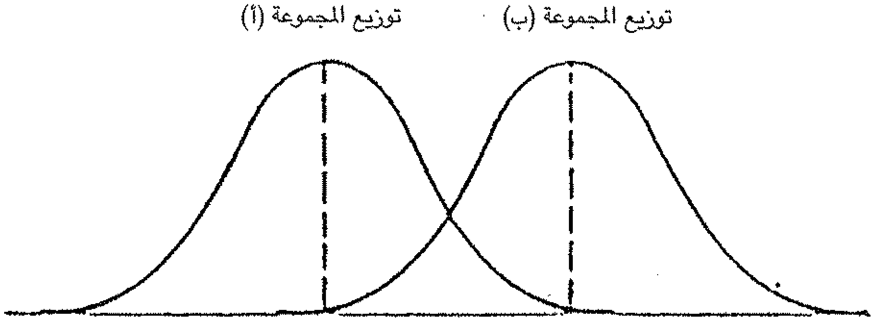
درجة معيارية اخرى، مشهورة خاصة لمستخدمي الاختبارات التحصيلية المقننة، تدريج مؤلف من ارقام بمنزلة واحدة تتراوح بين 1 الى9، وهذا التدريج يوفر طريقة لوصف اداء المفحوصين بشكل موسع. وتستخدم مثل هذه الدرجات في مدارس المقاطعات ولإيصال نتائج الإختبار للوالدين، اذ ان تدريج التسعة ذو بساطة واضحة ويستثنى الميل للتفسير الزائد للفروقات الصغيرة التي تظهر احياناً عندما تدون الرتب المئنية وحدها. ووسط هذا التوزيع = 5 نقاط، وانحرافه المعياري = نقطتين تقريباً. ونموذجياً في التوزيع الطبيعي فإن الدرجات الخام المناظرة لأوسط (20%) من التوزيع (اي من الرتبة 40 الى 59) يؤشر لهم بدرجة التساعي الاوسط (5) والدرجات الخام المناظرة للرتب المئينية السبعة عشر التالية (اي من رتبة 60 الى 76) يؤشر لهم بالتساعي السادس، والدرجات الخام المناظرة للرتب المئينية الاثنتي عشر التالية يؤشر لها بالتساعي السابع، والدرجات الخام المناظرة للرتب الميئنية السبعة التالية يؤشر لهم بالتساعي الثامن، والدرجات الخام المناظرة لاعلى أربعة رتب مئينية يؤشر لها بالتساعي التاسع (انظر شكل 19-2) والطريقة نفسها تستخدم في النصف الاسفل من التوزيع للتأشير على التساعيات الرابع والثالث والثاني والاول على التوالي.

الدرجات المرتبة في مستويات Scaled scores

معظم بطاريات الاختبارات التحصيلية تكون مرتبة (منظمة) في مستويات بحيث تزداد صعوبة المستويات بارتفاع مستوى الصف. ومع ذلك فإن المستوى الواحد من الاختبار يكون عادة مناسباً لصفين او اكثر من الصفوف المتتالية، فعلى سبيل المثال: المستوى الواحد قد يستهدف الطلبة من بداية الصف الثاني وحتى نهاية الصف الثالث. افترض ان طلبة الصف الثاني الذين درجاتهم على اختبار القراءة التحصيلي عند الرتبة المئينية (50) مقارنة بمجموعة الصف الثاني المعيارية، ورتبة الصف الثالث على الاختبار نفسه عند الرتبة المئينية (50) مقارنة بمجموعة الصف الثالث المعيارية ومن الواضح ان فئتي المفحوصين لا يكون لهم مستوى التحصيل القرائي المطلق نفسه، كذلك فان درجات مستوى الصف المعيارية لا تحقق هذا. وفي مثل هذه الاختبارات يكون الاهتمام. هو تحويل الدرجات بحيث تسمح بمقارنات معيارية ضمن كل مجموعة ويسمح بالوقت نفسه للتوزيع بان يقع على تدريج موحد يعكس المستوى التربوي او التطوري الاعلى للمجموعة الاكثر تقدماً (انظر شكل 19-3)، ومثل هذه الدرجات المعيارية التي تعكس كلا من مجموعة الرفاق والمجموعات (الحالات) الطولية تعرف على انها درجات مدرجة (مرتبة في مستويات scaled scores) أو تدريج المستوى الصفي.

إن بناء مثل هذا التدرج الذي يعكس ترتيب المستويات ظهر في النصف الاخير من القرن

السابق. واحدى الطرائق المبكرة هي طريقة ثيرستون في التدريج المطلق وكما تطبق على المشكلة الحالية فإن تدريج ثيرستون المطلق يعد طريقة او اسلوب لتحويل الدرجات الخام لكلا المجموعتين اللتين تقدمتا للاختبار نفسه، ومعروف انهما مختلفتان في المستوى التربوي او الحالة التطورية الى تدريج يكون توزيع درجات كل مجموعة فيه توزيع طبيعي، والدرجات على التدريج الجديد ستظهر كما في الشكل (19-3). لاحظ ان الهدف الرئيسي ليس ببساطة وضع الدرجات على التدريج نفسه. وحيث ان كلا المجموعتين تقدمتا للاختبار نفسه فانه سيتم التعبير عن الدرجات الخام على التدريج نفسه. ولنبدأ بالهدف وهو معايرة كلا التوزيعين بحيث يمكن اجراء مقارنات معيارية ضمن المجموعة وفي الوقت نفسه الاحتفاظ بالتدريج المشترك لدرجات المجموعتين. ووصف جوليكسن (1950). العملية الآتية لتطوير تدريج الدرجات المرتبة في مستويات:



شكل(19-3) توزيعات الدرجات المعيارية للمجموعتين أ و ب على تدريج مشترك .

1- تطبيق الاختبار على مجموعتين (أ و ب) شريطة ان يتداخل توزيع درجاتهم الخام .

2- اختر درجات خام تتراوح من 10 الى 20 (x_i) من منطقة التداخل .

3- لكل قيمة (x_i) احسب قيمتي درجات Z المعيارية المناظرة لـ Z_{bi} و Z_{ai} كل في توزيعه.

4- مثل بالرسم البياني ازواج الدرجات Z_{ai} و Z_{bi} لكل قيم درجات x_i الخام، بحيث تكون Z_{ai} على المحور الرأسي و Z_{b2} على المحور الافقي.

5- ان كان هذا الرسم البياني خطي، فان كلا المجموعتين يمكن معايرتها على تدريج الدرجات نفسه .

وعلى افتراض ان التمثيل حقق متطلبات الخطية، سنبين الآن كيف يمكن تأسيس تدريج مشترك لهذه الدرجات وربطها بتوزيعات درجات كلا المجموعتين .

دعنا نرمز الى التدريج المشترك على أنه Y، ولاحظ ان كل من تدريجYو Z_a هي تحويلات خطية كل منها للاخرى. اي ان اي قيمة لـ Z_{ai} على تدريج Z_a يمكن ربطها بتدريج Y بالتحويل الخطي:

$$\mu_{ya} + \sigma_{ya}(Z_{ai}) = Y_i \qquad (12\text{-}19)$$

وحيث ترمز μ_{ya} و σ_{ya} الى المتوسط والانحراف المعياري لدرجات Y للمجموعة A. والآن يمكن ربط كل قيمة y_i بتوزيع Z_b بالعلاقة :

$$\mu_{yb} + \sigma_{yb}(Z_{bi}) = Y_i \qquad (13\text{-}19)$$

لذا فإن:

$$(Z_{bi})\ \sigma_{yb} + \mu_{yb} = (Z_{ai})\ \sigma_{ya} + \mu_{ya}$$

ويمكننا حل التعبير هذا لـZ_{ai} على النحو التالي:

$$\frac{\mu_{yb} - \mu_{ya}}{\sigma_{ya}} + Z_{bi}\ \frac{\sigma_{yb}}{\sigma_{ya}} = Z_{ai} \qquad (14\text{-}19)$$

ويمكن وصف المحاور الاساسية للازواج المرتبة (Z_{ai}. Z_{bi}) بالمعادلة الخطية $(Z_{bi})\ k + m = Z_{ai}$. ويمكن حساب قيم k (الميل) وμ (القاطع) بطرائق مبينة في الفصل 16 (انظر المعادلات 16 –2أ و 16 –2ب)، وبجعل $\sigma_{y6} / (\mu_{ya} - \mu_{yb}) = m$ و $\frac{\sigma_{yb}}{\sigma_{ya}} = k$ وتأشير القيم مناسبة لكل من μ_{ya} و σ_{ya} يمكننا بعدها ايجاد قيم μ_{yb} و σ_{yb} لمعادلات القاطع والميل على التوالي. وحالما نحصل على قيم كل من μ_{ya} و σ_{ya} و μ_{yb} و σ_{yb} يمكننا تحويل الدرجات الخام على اي من التوزيعين A أو B إلى التدريج المشترك Y. ومع ان طريقة ثيرستون تسمح باختلاف كلاً من المتوسط والانحراف المعياري لكلا التوزيعين، فقد لاحظ جوليكس انه يبقى من الضروري افتراض ان كلا التوزيعين متشابهين في الشكل.

وفي مشكلة التدريج المذكورة اعلاه للمجموعتين المعروف اختلافهما في المستوى التربوي او الحالة التطورية التي تتقدم للمستوى الاختباري نفسه، فإن هدف التحليل هو تقنين توزيعات درجات المجموعتين على التدريج نفسه، مشكلة مشابهة ولكنها اكثر تعقيداً تكون عندما يقصد استخدام مستويين اثنين مع مجموعات مختلفة، ونريد التعبير عن درجات المستويين على التدريج نفسه، ومعايرة الدرجات للمجموعتين على التدريج. ثانية فإن الهدف يكون في استخدام التوزيعات المعيارية لإجراء مقارنات معيارية ضمن المستوى واستخدام التدريج المشترك بحيث يمكن مقارنة الاداء على المستويات المختلفة. ولحل هذه المشكلة يلزم اجراء خطوتين:–

الاولى : يجب وضع الدرجات من المستويين على التدريج نفسه، وتعد هذه مشكلة المعادلة الرأسية التي ستناقش في الفصل التالي.

الثانية : يجب تحويل الدرجات الناتجة عن المعادلة الرأسية بحيث يتم معايرة توزيع الدرجات الجديدة لكلا المجموعتين، وهذه الخطوة يمكن تحقيقها باستخدام طريقة التدريج المطلق التي بيناها سابقاً.

وكهدف في تفسير الدرجات المرتبة في مستويات في الاختبارات المقننة، يجب على المستخدمين ان يعرفوا أن مدى الدرجات المرتبة في مستويات يقع بين (1 و 100) أو (1و 1000). فعلى سبيل المثال في الاختبار الفرعي في القراءة في اختبارات الميترو بولتيان التحصيلية (Prescottetal, 1978) كانت درجة المفحوص الذي رتبته الميئنية في الصف الثاني (50)، كانت درجته على هذا التدريج (571) ومفحوص آخر في الصف الثالث له المكان النسبي نفسه كانت درجته على هذا التدريج (644) وفي الموقع نفسه. ولمفحوص في الصف الرابع كانت درجته على هذا التدريج (674)، لذا فإن المفحوص الذي يعمل من خلال التطور التربوي الطبيعي سيتقدم في كل سنة تالية لاختبار اكثر صعوبة ويحصل على درجة اعلى على هذا التدريج، وذلك لانه ببساطة له الموقع النسبي نفسه بالنسبة للاخرين في الصف نفسه او المستوى العمري .كذلك قد يتقدم مفحوصين اثنين لمستويات مختلفة ويحصلون على الدرجة نفسها على هذا التدريج فيما لو اجاب احد المفحصوين على عدد كبير من الفقرات السهلة اجابة صحيحة في اختبار المستوى السهل وفي الوقت نفسه اجاب الاخر عن عدد اقل من الفقرات اجابة صحيحة، ولكن تقدم لاختبار من مستوى اعلى، وتفيد الدرجات المرتبة في مستويات في تحقيق هدفين هما:

1- عندما يطبق على المفحوصين أختبار اعلى في المستوى من الصف الحالي الفعلي أو أدنى فإن درجاتهم على هذا التدريج يمكن استخدامها في ربط أدائهم بدرجات الرتب المئينية للمجموعة المعيارية او مستوى الصف الفعلي.

2- في الابحاث ودراسات التقويم حيث يحدد متوسط أداء المجموعة، وذلك للمجموعات التي يتقدم فيها المفحوصين الى مستويات اختبارية مختلفة، يمكن حساب متوسط هذا التدريج في الوقت الذي يكون متوسط الدرجات الخام أو متوسط الرتب الميئنية مضللاً.ومن المهم تذكر ان مع اهمية الدرجات على هذا التدريج في مقارنة اداء المفحوصين عند مستويات مختلفة ضمن الموضوع الواحد، الا انها غير قابلة للمقارنة عبر مجالات الموضوعات المختلفة ضمن بطارية الاختبار نفسه، ويمكن توضح هذه الحقيقة من خلال معايير اختبار الميتروبولتيان التحصيلي إذ حصل طالب في الصف

الثالث على درجة الرتبة المئينية (50) في كل من اختبارات القراءة والرياضيات، وحصل على درجات على هذا التدريج (644 و532) على التوالي في هذه الموضوعات، ولا يجب تفسير هذه الدرجات لتعني ان اداء المفحوصين في الرياضيات اضعف من ادائهم في القراءة بالنسبة لاقرانهم. والفروق في درجات التدريج هنا تعكس الفروقات في الصعوبة النسبية لإختبارات القراءة والرياضيات للمجموعة المعيارية لهذا المستوى الصفي. وتحديداًفإن المجموعة المعيارية للصف الثالث ككل كان اداءها افضل نوعاً ما في المادة المغطاة بإختبار القراءة اكثر مما هو في المادة المغطاة في اختبار الرياضيات بالنسبة لما سيكون عليه اداء المجموعات المعيارية في المستويات الصفية الاخرى.

مكافئات الصف والعمر:

عندما يختبر الاطفال بمقاييس الاستعداد او التحصيل فإن مستخدم الاختبار يحتاج الى درجات معيارية تشير الى كيفية أداء طفل مقارنة بأداء الاطفال الاخرين من عمر معين او مستوى صفي معين. وتستخدم احياناً لهذه المقارنات درجات المكافئات الصفية أو العمرية. ومع ذلك توجد محددات شديدة لهذه الدرجات والتي تصبح واضحة حال وصف عملية تطويرها.

وقد وصفت اشترناكت (Echternacht, 1977) العملية العامة، التي يبني بواسطتها ناشري الاختبار درجات المكافئات الصفية، وهذه العملية تتبع الخطوات الآتية:

1- تحويل توزيع درجات الاختبار الخام الى توزيع درجات مرتبة في مستويات (نوقشت في الجزء السابق) وتكرر هذه العملية لعدة مستويات صفية متتالية.

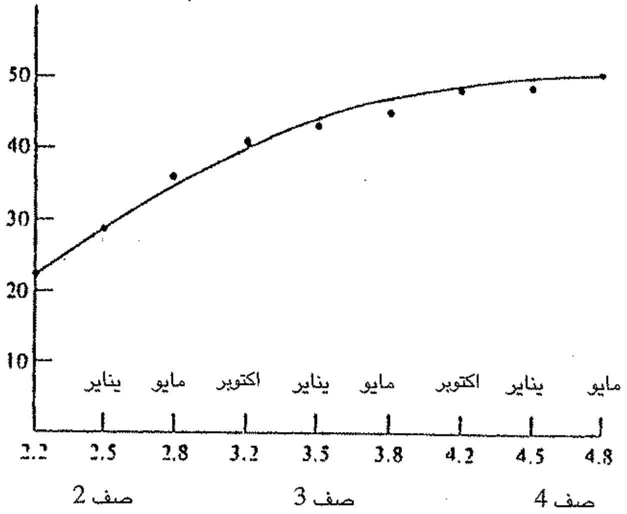
2- تحديد وسيط درجة التدريج لكل مستوى صفي، وتمثل على مستوى ثنائي المتغير، ويشير المحور الافقي الى الصف والسنة والشهر الذي أجريت فيه العملية الاختبارية ويشير المحور العمودي الى قيم الدرجات المرتبة في مستويات، ويكون وسيط التدريج للمستوى الصفي اعلى السنة والشهر لليوم الذي تم فيه معايرة الاختبار .

3- رسم منحنى املس يوصل ما بين هذه النقاط كما هو مبين في الشكل (19-4) (بشكل نموذجي تكون العلاقة منحنية وليست خطية).

وحالما يتم معرفة درجة المفحوص الخام فإنه يمكن تحويلها الى درجة التدريج ثم الى درجة تدريج المستوى الصفي من خلال هذا المنحنى (او القيم المجدولة التي نتجت عن هذا المنحنى) وتسمح هذه الطريقة بعمل امتدادات من الدرجات الخام لاختبار معين الى المستويات الصفية التي لم يطبق عليها الاختبار ويسمح ايضاً بتحويل الدرجة الخام الى امتدادات القيم ضمن المستوى الصفي في أوقات غير تلك التي اجري فيها الاختبار وكما

سنرى فإن هذه العملية تفتح باباً لبعض النقد لدرجات المكافىء الصفي. وأكثر محددات درجات المكافىء الصفي هي:

1- تكون المكافئات العمرية أكثر عرضة لخطأ التفسير من قبل مستخدمي الاختبار أكثر من الانواع الاخرى من الدرجات المعيارية. افترض على سبيل المثال أنه اختبر طالب في الصف الثاني في نهاية السنة وحصل على مكافىء صفي G.E =5.3 في كل من الرياضيات والقراءة والعلوم والتفسير المناسب هو انه هذه الاختبارات ولمحتوى الصف الثاني فإنه يتوقع تشابه أداء هذا الطالب مع أداء تلميذ عمره خمسة سنوات وثلاثة اشهر، وهذا لا يعني ان أداء هذا المفحوص سيكون بالطريقة نفسها على اختبار يغطي محتوى الصف الخامس، وذالك لأن هذا التلميذ لم يتعرض للتدريس لتلك المادة، لذا فإن اسم هذه الدرجة يعطي انطباع خاطىء مؤداه ان الطالب له قدرة على انجاز مهام الصف الخامس.



شكل(19-4): تمثيل بياني لقيم درجات خام مختارة لمستويات صفية متتالية في مواقف اختبارية ثلاثة

2- مقارنات درجات المستوى الصفي عبر مجالات عدة للمفحوص نفسه قد تكون مضللة أيضاً. على سبيل المثال مكافىء صفي (3.9) في القراءة و(3.9) في العلوم يبدو انها تشير لكفاءة متساوية في كلا المجالين، ولكن هذا ليس حقيقياً. مع ان المكافئات الصفية لكلا الموضوعين متساوية، الا ان الرتب المئينية في المجموعة المعيارية للمستوى الصفي للموضوعين قد تكون (68) في القراءة و(82) في العلوم .

3- معدل النمو على تدريج المكافئات الصفية ضمن مستوى صفي معين او مستوى عمري حسبت بوساطة عدد قليل من النقاط المناظرة للفترات الزمنية التي طبقت فيها الاختبارات. والمنحنى الواصل بين هذه النقاط يرسم عادة بإفتراض ان معدل النمو او التطور يكون منتظم بين هذه النقاط ومستمر، وفي موضوع مثل القراءة على سبيل المثال فإن الافتراض للكميات نفسها من التطور من شهر لشهر خلال السنة الدراسية قد يكون متعذر. كذلك فيما لو حصل طالبين على درجات (3.8) و(3.9) على التوالي، فليس من الضرورة ان يكون صحيحاً اعطاء شهر تعليمي اضافي للتلميذ الاول ان يجعل التلميذين الاثنين بالمستوى الوظيفي نفسه.

4- للاختبار الواحد، والمناسب للمفحوصين لمدى او فئة محددة من المستويات الصفية (مثل من الثالث وحتى الخامس)، تحول الدرجات الخام عند طرفي التوزيع الى تدريج المكافيء الصفي بالاعتماد على امتداد المنحنى لمجموعات المستوى الصفي والذين لم يتقدموا للاختبار فعلاً، فعلى سبيل المثال من الممكن لطلبة الصف الخامس ان يتقدموا لاختبار في الدراسات الاجتماعية للمستوى الصفي نفسه والحصول على درجة مكافيء صفي =1.0 ، ومع ان اختبار الصف الخامس لم يطبق على الذين سيدخلون مستوى الصف الاول وتم الحصول على امتدادات درجات المكافيء الصفي من خلال افتراض ان المنحنى او الخط الذي يصف العلاقة بين درجات الاختبار والمستوى الصفي يستمر في التنبؤ للصفوف التي لم تختبر هو افتراض معرض للمساءلة. لذا فإن استخدام درجات المكافئ الصفي تكون اكثر عرضة للمساءلة على طرفي التوزيع. وقد طورت تدريجات المكافيء العمري بالطريقة العامة نفسها التي استخدمت في تطوير تدريجات المكافيء الصفي. ويجب تأسيس قرار محكي في تدريج المكافيء العمري وذلك لتعيين مفحوصين للمجموعات العمرية الخاصة، فلكل طفل عمره (سنتان أو أقل) يمكن ان تعطى فئات العمر على فترات من ثلاثة أشهر أو ستة أشهر. وقد تكون الفئات العمرية للاطفال الرضع بفترات اسبوعية أو شهرية. وللمفحوصين الاكبر فإن الفترات السنوية تعد معياراً مناسباً. ومن الشائع في تعيين المفحوصين الى الفترات السنوية تعيين المفحوصين الذين هم ضمن ستة اشهر تالية لتاريخ الولادة إلى فئة العمر الاعلى من فئتهم العمرية، ومن الممكن ايضاً استخدام المفحوصين الذين هم ضمن ثلاثة او اربعة اشهر تالية لتاريخ الولادة، وبتعيين المجموعة الاخيرة الى الفئة التالية الاكبر عمراً وهذا ينتج عنه خسارة افراد من المجموعة المعيارية للفئة العمرية .

الخلاصة:

تعد الدرجات المعيارية مهمة عندما يريد مستخدم الاختبار مقارنة درجة المفحوص بتوزيع درجات عينة مأخوذة بطريقة محددة تماماً من المجتمع. ويسمح اختيار العينة المعيارية عبر المعاينة الاحتمالية لمطور الاختبار ان يحسب درجة خطأ المعاينة بإحصاءات العينة وخفض فرصة التحيز المنتظم في البيانات المعيارية. وقد اقترحت تسع خطوات يمكن اتباعها في اجراء دراسة المعايرة وعندما يحدد مطور الاختبار كمية الخطأ المسموح به في متوسط عينة المعايرة يمكن استخدام هذا الخطأ المعياري في حساب اصغر حجم عينة يلزم للمعايرة.

وتوجد اربعة انواع من المعانية الاحتمالية التي يمكن استخدامها في دراسات المعايرة وهي:

المعاينة العشوائية البسيطة، والمعاينة المنتظمة، والمعاينة العشوائية العنقودية، والمعاينة الطبقية. وكقاعدة عامة فإن المعاينة العشوائية العنقودية تؤدي الى تقديرات اكثر دقة للعينة عندما يرتبط الاداء على المتغير العنقودي مع المتغير الذي نهتم به. وتستخدم المعاينة الطبقية عادة معايرة الاختبارات المعيارية لاسباب الملائمة التي ينتج عنها اكبر اخطاء معاينة، وبالتالي يلزم اختبار أعداد اكبر من المفحوصين للاحتفاظ بدرجة الخطأ الناتج عن العملية الاختبارية لعينة عشوائية بسيطة أصغر، والتي قد تستخدم أنواع مختلفة من المعاينة عند مراحل مختلفة. وأحد النماذج الشائعة هو اختيار عينات تجمع في المؤسسات يتبعها معاينة عشوائية بسيطة او عنقودية لافراد كل طبقة او تجمع .

وفي وصف دراسة المعايرة في دليل الاختبار على الفاحص توثيق طريقة المعاينة المستخدمة، ومعدل الرفض أو رفض الاستجابات من قبل الذين تمت دعوتهم للاشتراك، ويوم الدراسة، والخصائص الديموغرافية للعينة ويجب ان يرافق إحصائيات العينة على الاقل قيمة الخطأ المعياري للمتوسط .

ولتدعيم القيمة التفسيرية لدرجات الاختبار يجب تطوير عدد من التحويلات المعيارية للدرجات والتي تفيد في المقارنة بين المفحوصين. وفي هذا الفصل تم وصف انواع الدرجات المعيارية الآتية :

1- **الرتب المئينية** : وتعكس النسبة المئوية للمفحوصين في المجموعة المعيارية الذين درجاتهم تساوي درجة خام معينة أو تقل عنها .

2- **الدرجات الزائية الخطية** : وتحدد على انها نسبة انحراف درجة المفحوص الى الانحراف المعياري للمجموعة.

3- **الدرجات الزائية المعيارية** : وهي القيم الزائية المستخرجة من الجدول الزائي المعياري والمناظر لرتبة معينة.

4- **الدرجات المشتقة**: وهي تحويلات خطية للدرجات الزائية (بما فيها الدرجة التائية، ودرجة نسبة الذكاء المشتقة، ودرجات NCE).

5- **التساعيات** : وتنشأ من تقسيم تدريج الدرجات الخام الى تسعة اجزاء بحيث تقع نسبة معينة من الدرجات في كل جزء.

6- **الدرجات المرتبة في المستويات**: وتعكس درجة المفحوص بالنسبة للمجموعة المعيارية وموقع توزيع المجموعة المعيارية بالنسبة لتوزيعات المجموعات الاخرى.

7- **المكافئات الصفية والعمرية**: وتؤشر مستوى المفحوص النموذجي في مستوى عمري أو صفي معين على الاختبار.

ويجب التدريب على الحذر في تفسير كل نوع من الدرجات المعيارية. ومن المرغوب به عرض البيانات التي تبين كيفية تأشير كل نوع من الدرجات المعيارية بخطأ القياس عند نقاط مختلفة على توزيع الدرجات.

التمارين:

1/ بتتبع التوزيع الافتراضي لمجتمع مؤلف من (50) مفحوص لاختبار حل مشكلات غير لفظي لمفحوصين أُخذوا من خمسة صفوف منفصلة في مدرسة واحدة. افترض ان مستخدم الاختبار يرغب في تطوير معايير صفية محلية لهذا الاختبار، وحيث ان التطبيق كان فردياً فإن المجتمع الكلي لم يتم اختباره .

أ- انشيء عينة عشوائية بسيطة مؤلفة من (20) مفحوص من هذا المجتمع. احسب الخطأ المعياري للمتوسط.

ب- بإستخدام الصفوف على أنها الطبقات، اعد عينة طبقية تتألف من 20 مفحوص. إحسب الخطأ المعياري للمتوسط .

ج- استخدم الافراد المفحوصين المنخرطين في برامج الطلبة غيرالعاديين على انه المتغير الطبقي. فهل توزيع هذا المتغير في العينة التي اعددتها يعكس توزيعها في المجتمع .

د- قارن وسط عينتك بوسط المجتمع الكلي. وهل دقة وسط عينتك يعكس ما يمكن التنبؤ به بإستخدام نظرية المعاينة.

هـ- قارن نتائجك بنتائج الاخرين في الفصل. ومن خلال نتائج الفصل، أعد توزيع المتوسطات الناتجة عن طرائق المعاينة الثلاثة. أي هذه الطرائق يبدو انها تؤدي الى تقديرات اكثر دقة.

الفصل 1	الفصل 2	الفصل 3	الفصل 4	الفصل 5
101	*96	103	*93	104
102	97	100	91	x96
*100	96	101	*88	100
99	95	104	*90	101
*98	*91	*107	92	101
100	97	102	* 89	*99
97	*93	105	87	*99
101	96	*106	*88	92
99	94	99	92	97
*97	*95	*103	90	101

* الطلبة غير العادين:

2/ لكل من الطلبة ادناه، وصفت طريقة المعاينة اقرأ كل وصف، ثم حدد نوع المعاينة الأحتمالية المستخدمة.

أ- استاذ علم نفس ربط بين استبانة مختصرة تتعلق بعادات الطلبة الدراسية، وذلك لكل من صيغ الاختبار الاربعة، ثم وزع الاختبار على الطلبة عندما حضروا غرفة الصف.

ب- مشرف على الخدمات الطلابية في الجامعة اختار عشوائياً خمس وحدات من مخيم، ومن نزلائها اختار عينة عشوائية تألفت من 60% من النزلاء الجدد و40% من الفئات الاخرى وذلك في دراسة مسحية تتعلق بالادمان الكحولي.

ج- اختير العاملين في شركة كبيرة في دراسة صدق، وذلك اذا كانت الارقام الثلاثة

الاخيرة من الرقم المبين في بطاقة التأمين الاجتماعي تظهر نفسها في جدول الارقام العشوائية

3/ متخصص في التربية الصحية صمم اختبار تألف من (30) فقرة لتقييم المعرفة في الحقائق الصحية الاساسية، ويريد هذا المختص معايرة الاختبار على عينة طلبة الصف التاسع في العلوم الصحية ويرغب في ان يتأكد بنسبة 95% بإن وسط المجتمع يقع ضمن فترة نقطة واحدة أعلى أو أدنى من تقديرالعينة. ويرغب هذا المختص بإن يكون وسط للاختبار المؤلف من 30 فقرة (25) درجة تقريباً وبإنحراف معياري (5) تقريباً. فما اصغر حجم عينة يحقق مستوى الدقة المرغوب به في هذه المعايرة .

4/ أ- اكمل جدول تحويل الدرجات المعيارية المبينة في الجدول أدناه:

X	التكرار	الدرجة الخام	الرتبة المئينية	درجة Z الخطية	درجة Z المعيارية	الرتبة المئينية المعيارية
45	3					
44	7					
43	8					
42	15					
41	20					
40	15					
39	13					
38	9					
37	7					
36	3					

ب- مثل بيانياً النقاط التي تبين العلاقة بين الدرجات الخام والرتب المئينية المعيارية (بحيث تكون الدرجات الخام على المحور الافقي والرتب المئينية على المحور العمودي)، أرسم المنحنى الاملس الاكثر توافقاً مع هذه البيانات .

ج- الان وعلى المحاور نفسها مثل بيانياً النقاط التي تبين العلاقة بين الدرجات الخام والرتب المئينية الاصلية، ارسم المنحنى الاملس الاكثر توافقاً مع هذه البيانات. فسر الفروق في شكل كلا المنحنيين .

5/ اقرأ وصف كل من المواقف الآتية واجب عن الاسئلة التي تلي كل منها والذي يتطلب تفسير التدريجات المعيارية المختلفة.

أ- الرتبة الميئنية لكارل (98) ولكاثي (53) وذلك في اختبار تحصيلي مقنن. وعند اعادة الاختبار ازدادت كل درجة بقيمة 2 درجة خام فاي المفحوصين كان كسبه أكبر في الرتبة المئينية؟ لماذا؟

ب- كان ترتيب جوي على التساعي الثاني في السنة السابقة وعلى التساعي الثالث في السنة الحالية وذلك قي اختبار الحساب الرياضي، بينما كان ترتيب لويس على التساعي الرابع في كلا السنتين. فهل هذا يعني ان جوي تقدم بشكل افضل من لويس بالنسبة للمجموعة المعيارية؟ لماذا؟

و- بعد اعادة الاختبار نفسه هبطت درجة لاري من الرتبة المئينية (49) الى الرتبة المئينية (45)، بينما هبطت درجة جون من الرتبة الميئنية (89) الى الرتبة المئينية (85). اين يكون هبوط الدرجات الخام مكافئاً لهذا الهبوط في الدرجات المئينية هذه؟ لماذا؟

د- في نهاية الصف الثالث كانت المكافئات الصفية لكل من لويس، ساندرا، كيم على النحو الآتي: 3.2 ، 3.5، 3.8، فهل لهؤلاء المفحوصين مسافات متساوية بين درجاتهم الخام؟ بين رتبهم المئينية؟ وبين درجاتهم المرتبة في مستويات؟

6/ على افتراض ان الجداول المعيارية لاختبار تحصيلي مقنن تعتمد تطبيق الاختبارات في الشهر التاسع لكل مستوى صفي. والقيم المبينة في الجدول ادناه هي درجات الوسيط المرتبة في مستويات ناتجة عن اختبار طلبة كل مستوى صفي:

الصف	الدرجة المرتبة في القراءة	الدرجة المرتبة في العلوم
0.9	375	325
1.9	542	420
2.9	637	492
3.9	680	548
4.9	700	605
5.9	715	640

أ- استخدم هذه البيانات في رسم منحنى يبين العلاقة بين المستوى الصفي والدرجات المرتبة في مستويات لكل إختبار .

ب- اعرض التفسيرات الممكنة للفرق في شكل كلا المنحنيين .

ج- من خلال المنحنى، حدد قيم المكافيء الصفي المناظر تقريباً للدرجات المرتبة في مستويات لكل إختبار .

د- لو تم اختبار مفحوصين من الشهر الثالث للفصول (1، 2، 3) فبماذا تتنبأ عن وسيط الدرجات المرتبة في المستويات.

هـ- افترض أن درجات الصف السادس كانت أدنى بصف في العلوم، وطلبة الصف الثاني كانت درجاتهم اقل بصف بالمبحث نفسه. فهل هؤلاء التلاميذ يحتمل أنهم مشابهون في موقعهم للمجموعة الصفية المعيارية؟ وهل ان درجات إختلافهم متساوية في الجدية؟

و- من خلال المعايير الصفية، هل المكافئ الصفي الواحد دون المستوى الصفي متكافئ في جديته في القراءة كما في العلوم.

7/ باستخدام الدرجات الزائية المعيارية المحسوبة في السؤال الرابع أعد جدول لتحويل الدرجات الخام في ذلك التوزيع الى الدرجات التائية، وإلى درجات نسبة الذكاء الانحرافية وإلى الدرجات على تدريج (ETS) نفسه المدون لدرجات الاختبارات الفرعية في (GRE) أو (SAT).

الفصل العشرون

20

معادلة درجات الاختبارات المختلفة

الفصل العشرون

معادلة درجات الاختبارات المختلفة

في عديد من المواقف يقاس مفحوصين مختلفين بأدوات (اختبارات ومقاييس مختلفة) يفترض أنها تقيس السمة نفسها. فمثلاً في برامج شهادات الاختصاص هنالك صيغ عدة للاختبار تبنى بحيث يتعاقب استخدامها لمرات عديدة خلال السنة الواحدة. فإن كان توزيع الدرجات لهذه الصيغ ليس نفسه فمن الضروري تحديد الدرجات المتكافئة من الصيغ الاختبارية المختلفة، ويطلق على عملية تحديد الدرجات المتكافئة للصيغ المختلفة اسم المعادلة الأفقية "Horizontal equating". أيضاً تتكون بطارية الاختبارات التحصيلية من عدة مستويات كل مستوى يناسب طلبة مدى معين من الصفوف الدراسية. افترض أن أحد هذه المستويات في البطارية مناسب لطلبة الصف الأول من المرحلة الثانية (الأول المتوسط) والمستوى الثاني يناسب طلبة نهاية المرحلة المتوسطة، وفي مثل هذه الاختبارات فإن الطلبة الذين يتقدمون ببطء في المرحلة المتوسطة يُعد اختبار المستوى الثاني صعب جداً للقياس الدقيق لهذه الفئة من الطلبة، والحل الممكن لهذه الفئة هو تطبيق اختبار المستوى الأول. ولتفسير أدائهم من الضروري تحديد الدرجات المتكافئة لهم في المستويين، ويطلق على هذه العملية اسم المعادلة العمودية "Vertical equating". وأخيراً افترض أن باحثاً أجرى دراسة استنبط متغيراتها من ملاحظات المعلمين وترتبط متغيراتها بدرجات اختبارات التحصيل المعيارية، وإذا طبق الباحث بطاريات اختبار مختلفة في المقاطعات المختلفة، فمن الضروري معادلة درجات الاختبارات المختلفة، ويعد هذا مثالاً على المعادلة الأفقية للدرجات.

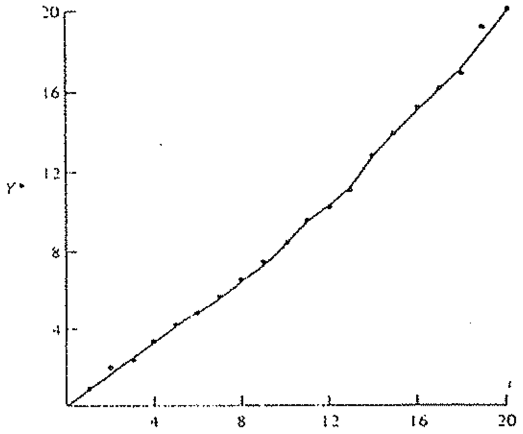
ويمكن اعتبار عملية المعادلة تأسيس أو اشتقاق درجات متكافئة من اختبارات مختلفة. فإذا عودلت درجات اختبارين (X وY) وتقدم مفحوص لاختبار (x) فإنه يمكننا تحديد الدرجة المكافئة لها على الاختبار (y). والدالة التي تربط بين درجات (x) والدرجات المكافئة لها على اختبار (y) يمكن كتابتها على النحو الآتي:

$$f(x) = y^*$$

ويمكن مقارنة درجة المفحوص الذي تقدم لاختبار (x) بدرجة مفحوص آخر تقدم لاختبار y، وذلك بتحويل درجات (x) إلى تدريج (y^*). ومشكلة المعادلة تكمن في تحديد الدالة f (x). ويمكن تمثيل الدالة بمعادلة رياضية على نحو:

$$a(x - c) + d = y \quad \text{(1-20)}$$

أو يمكن تمثيلها بيانياً كما في الشكل(20-1)، والنقطة على محور (Y) المناظرة للدرجة على محور (x) هي (*y) والمكافئة للدرجة (x).



شكل (20-1): تمثيل بياني لمعادلة الرتب المئينية لدرجات اختبارين افتراضين كل منهما مؤلف من 20 فقرة.

تعريف المعادلة:

كما وصفت أعلاه فهي عملية تكوين درجات متكافئة على اختبارين، ولم نتطرق للشروط الواجب توافرها لاعتبار الدرجات متكافئة. والتعريف الشائع والمقبول للمعادلة هو: يمكن اعتبار درجتين أحدهما على الاختبار (x) والأخرى على الاختبار (y) متكافئة إذا كانت كل منهما تقيس السمة نفسها بدرجات ثبات متساوية ورتب مئينية نفسها. ويعد هذا التعريف للدرجات المتكافئة الأساس لطرائق المعادلة الخطية والمئينية، والتي سيتم طرحها في الجزء التالي.

جمع البيانات اللازمة للمعادلة:

توجد ثلاثة تصميمات أساسية لجمع البيانات اللازمة لإجراء عمليات المعادلة. في التصميم الأول بتقدم مجموعات مفحوصين مختلفين للاختبارات المختلفة، ويتم اختيارها عشوائياً. فمثلاً لمعادلة صيغتي اختبار يقسم المفحوصين على مجموعتين تتقدم كل مجموعة إلى أحدى صيغتي الاختبار،ويطلق على هذه التصميم التصميم (أ). ويتقدم المفحوصين في حالة التصميم (ب) جميعهم لصيغتي الاختبار، ولضبط أثر ترتيب الاختبارات يتم تعريض المجموعات المختلفة للاختبارات بكل طرائق الترتيب الممكنة للاختبارات، وترتيب المفحوصين

يتم تحديده عشوائياً. وفي حالة معادلة صيغتي اختبار توزع المجموعة الكلية إلى مجموعتين عشوائياً، وتتقدم المجموعة الفرعية الأولى للاختبار x أولاً ثم تتبع بالاختبارY. وتتقدم المجموعة الأخرى للاختبارين بالترتيب المعاكس. وفي التصميم (ج) تقدم الاختبارات المختلفة إلى مجموعات مختلفة من المفحوصين بالإضافة لاختبار مشترك يتقدم إليه جميع المفحوصين ويكون هذا الاختبار على الأغلب أقصر من الاختبارات المراد معادلتها. ولمعادلة صيغتي اختبار تتقدم كل مجموعة إلى أحد الاختبارين بالإضافة إلى الاختبار المشترك، وليس بالضرورة في هذا التصميم توزيع الأفراد عشوائياً على المجموعتين. ويمثل الجدول(20-1) مخطط لهذه التصميمات.

جدول (20-1): مخطط توضيحي للتصميمات الثلاثة لمعادلة الاختبارات.

التصميم	المجموعة	
أ	X^1	Y^2
ب	(أ) Y:X	X:Y
جـ	(ب) Z:X	Z:Y

أ- أ: تشير إلى أن المجموعة تقدمت للاختبار x ثم للاختبار y .

ب- ب: تشير إلى أن المجموعة تقدمت للاختبار x ثم للاختبار المشترك z.

المعادلة الخطية للتصميم أ:

تذكر أن تصميم (أ) يتطلب توزيع عشوائي لمجموعتي الاختبار. وفي المثال الآتي نفترض أننا سنعادل درجات اختبارين، وفي هذه الحالة سيتم تشكيل مجموعتين عشوائياً تتقدم كل منهما لإحدى الصيغتين. وتعتمد المعادلة الخطية على الافتراض بأن الفروق في المتوسطات والانحراف المعياري وتوزيع الدرجات على الصيغتين X و Y متساوية.

وإن كان هذا صحيحاً فإنه يمكن تعريف الدرجات المتكافئة بتحديد أزواج من الدرجات كل منهما على إحدى الصيغتين X أو Y، ويكون لهما الدرجة الزائية نفسها. ولهذا تكون الدرجتين متكافئتين إذا كان:

$$\frac{y - \hat{\mu}_y}{\hat{\sigma}_y} = \frac{x - \hat{\mu}_x}{\hat{\sigma}_x}$$

ولهاتين الدرجتين الرتبة المئينية نفسها، ومعادلة التحويل من درجات X على Y هي:

$$a(x - c) + d = y^* \qquad (1\text{-}20)$$

وتكون القيم في هذه المعادلة:

$$\frac{\hat{\sigma}_y}{\hat{\sigma}_x} = a \qquad \text{2-20 أ}$$

$$\hat{\mu}_x = c \qquad \text{2-20 ب}$$

$$\hat{\mu}_y = d \qquad \text{2-20 جـ}$$

وتمثل x الدرجة على الصيغة الأولى و*y الدرجة على الصيغة المحوّلة، وتصبح الدرجة x على الاختبار الأول مكافئة للدرجة *y على الاختبار الثاني.

مثال: افترض أن المجموعة الأولى تقدمت للاختبار x وكان متوسطها = 50 وانحرافها المعياري = 10 في حين أن المجموعة الأخرى تقدمت للاختبارy كان متوسطها = 52 وانحرافها المعياري =11. بالتعويض بالمعادلة (20-2) نحصل على :

$$1.1 = \frac{11}{10} = a$$

$$50 = c$$

$52 = d$ ، لذا تصبح المعادلة (1-20) على النحو الآتي:

$$52 + (50 - x)\ 1.1 = Y^*$$

وعلى هذا فإن الدرجة 45 على الاختبار x تكافئ الدرجة 46.5 على تدريج درجات اختبار y.

وتعد المعادلة الخطية مناسبة فقط إذا كان توزيع درجات كل من x و y تختلف فقط في المتوسط والانحراف المعياري. والطريقة الأخرى للمعادلة هي معادلة الرتب المئينية المتساوية والتي تحدد الدرجات التي لها الرتبة المئينية نفسها على توزيع درجات كلا الاختبارين. وسيتم توضيح هذا في أجزاء لاحقة.

المعادلة الخطية للتصميم ب

هنا يتقدم المفحوصين للصيغ الاختبارية المختلفة ولكن بترتيب مختلف، وذلك للاحتمالات

الممكنة جميعها بحيث يتم تحديد الترتيب عشوائياً على المجموعات المختلفة. وفي حالة معادلة صيغتي اختبار هنالك احتمالين اثنين فقط، وتكون قيم كل من a و c و d على النحو الآتي:

$$a = \sqrt{\frac{\hat{\sigma}_{y}{}^{2}{}_{1} + \hat{\sigma}_{y}{}^{2}{}_{2}}{\sigma_{x}{}^{2}{}_{1} + \sigma_{x}{}^{2}{}_{1}}} \qquad \text{3-20 أ}$$

$$c = \frac{\hat{\mu}_{x1} + \hat{\mu}_{x2}}{2} \qquad \text{3-20 ب}$$

$$d = \frac{\hat{\mu}_{y1} + \hat{\mu}_{y2}}{2} \qquad \text{3-20 جـ}$$

وتمثل الأرقام الدليلية في هذه الصيغ إلى المجموعات الفرعية التي تم تشكيلها لتتقدم للاختبارات في كلا الترتيبين.

ويقدم الجدول (20-2) وصفاً لإحصائيات دراسة افتراضية أجريت على مجموعتين وفقاً للتصميم ب.

جدول (20-2): الإحصاء الوصفي لمعادلة افتراضية باستخدام التصميم ب.

المجموعة	الصيغة	μ	σ
1	X	25.2	5.3
	Y	26.1	5.0
2	X	24.8	5.6
	Y	26.5	4.9

وبالتعويض في المعادلة (20-3) نحصل على:

$$a = \sqrt{\frac{24.01 + 25.00}{31.36 + 28.09}} = 0.90$$

$$c = \frac{24.8 + 25.2}{2} = 25$$

$$d = \frac{26.5 + 26.1}{2} = 26.3$$

وبالتعويض ثانية بالمعادلة (20–1) نحصل على:

$$y^* = 26.3 + (25 - x)\ 0.90$$

ولتحديد قيمة *y المكافئـــة للدرجة x = 30، نجـد أن الدرجة المكافئـــة لها على تدريج y =30.8.

المعادلة الخطية للتصميم جـ:

يتقدم لكل اختبار بهذا لتصميم مجموعة مختلفة،وليس بالضرورة تشكيل المجموعات عشوائياً، بالإضافة إلى اختبار مشترك يتقدم إليه المفحوصين جميعهم. لهذا عند معادلة درجات اختبارين يتقدم لكل اختبار مجموعة مختلفة بالإضافة إلى الاختبار المشترك.

افترض أن z تمثل درجات الاختبار المشترك، وتقدمت المجموعة الأولى للاختبار x والمجموعة الثانية للاختبارy. ولوصف الافتراضات التي تبنى عليها المعادلة الخطية من الضروري تحديد المجموعة الفرعية 1 وهي المجموعة التي تم اختيار المجموعة الفرعية 1 منها، والمجموعة 2 وهي المجموعة التي تم اختيار المجموعة الفرعية 2 منها. وتمثل المجموعتين الفرعيتين 1، 2 المجتمع الكلي. والافتراضات التي تبني عليها المعادلة الخطية هي:

1. تساوي الميل والقاطع والخطأ المعياري لانحدار x على z في المجموعة الأولى مع الميل والقاطع والخطأ المعياري لانحدار x على z للمجتمع بأكمله.
2. تساوي الميل والقاطع والخطأ المعياري لانحدار y على z في المجموعة الثانية مع الميل والقاطع والخطأ المعياري لانحدار y على z للمجتمع بأكمله.

وتكون هذه الافتراضات منطقية فيما لو تم تكوين مجموعتين عشوائياً، وإن لم يكن توزيع المجموعتين عشوائياً فإن الافتراضات لا يمكن الدفاع عنها **(Angoff,1971)**. وقد بين انغوف أنه كلما كان الفرق بين المجموعتين على الاختبار Z أكبر فإن الالتزام بالافتراضات يكون أضعف. وكما في التصميم P للمعادلة الخطية نستخدم المعادلة (20-1) في المعادلة ولكنها تصبح:

$$a = \sqrt{\frac{\hat{\sigma}^2_{y2} + b^2_{yz2}(\sigma^2_{Z} - \sigma^2_{Z2})}{\hat{\sigma}^2_{X1} + b^2_{Xz1}(\sigma^2_{Z} - \sigma^2_{Z1})}} \qquad 4\text{-}20$$

$$\hat{\mu}_{x_1} + b_{xz_1}(\hat{\mu}_z - \hat{\mu}_{z_1}) = c \quad \text{4-20 ب}$$

$$\hat{\mu}_{y_2} + b_{yz_2}(\hat{\mu}_z - \hat{\mu}_{z_2}) = d \quad \text{4-20 جـ}$$

وترمز الأرقام الدليلية في هذه الصيغ إلى المجموعات الفرعية. وتحسب الإحصائيات الخالية من الأرقام الدليلية للمجموعة الكلية.

وتمثل الرموز b_{xz_1}، b_{xz_2} انحدار x على z في المجموعة الأولى وانحدار y على z في المجموعة الثانية على التوالي. ويجب أن يلاحظ القارئ أن المعادلة (20-4 ب) تستخدم العلاقة بين x و z للحصول على تقدير لمتوسط درجات المجموعة الكلية، كذلك فإن المعادلة (20-4 جـ) تؤدي إلى تقدير متوسطات درجات y للمجموعة الكلية. والبسط والمقام في المعادلة (20-4 أ) هي تباين درجات X، Y على التوالي وللمجموعة الكلية. ويبين الجدول (20-3) وصفاً لإحصائيات الدراسة لمعادلة افتراضية أجريت على وفق التصميم جـ.

جدول (20-3): احصائيات وصفية لمعادلة افتراضية باستخدام التصميم جـ.

المجموعة	الاحصائي	المتغير		
		X	y	Z
1	$\hat{\mu}$	77.8		78.8
	$\hat{\sigma}$	12.7		11.2
	b_{xz1}	0.9		
2	$\hat{\mu}$		83.5	83.4
	$\hat{\sigma}$		11.1	8.2
	b_{xz2}		1.2	
المجتمع باكمله	$\hat{\mu}$			81.8
	$\hat{\sigma}$			9.3

وبالتعويض بالمعادلة (20-4) نحصل على:

$$1.07 = \frac{(^{2}8.2 - ^{2}9.3)\ ^{2}1.2 + ^{2}11.1}{(^{2}11.2 - ^{2}9.3)\ ^{2}0.9 + ^{2}12.7} = a$$

$$80{,}5 = (78.8 - 81.8)\ 0.9 + 77.8 = c$$

$$81{,}48 = 1.2\ (83.4 - 81.8) + 83{,}4 = d$$

وبتعويض قيم a, c, d بالمعادلة (20-1) نحصل على:

$y^* = 1{,}1\ (x - 80.5) + 81{,}5$, وتستخدم في تحديد درجة y المكافئة للدرجة على اختبار x.

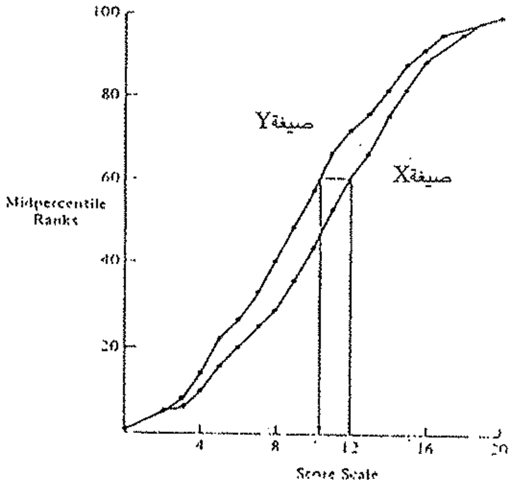
معادلة المئينات المتساوية : Equipercentile Equating

ذكرنا في السابق أنه يلزم لمعادلة المئينات المتساوية تحديد الدرجة التي لها الرتبة المئينية نفسها على الاختبارين، مع العلم أن طريقة المعادلة للتصميمين (أ,ب) هي نفسها، وسيتم شرحها على التوالي. بينما الطريقة المستخدمة للمعادلة للتصميم (جـ) معقدة ولن نعرضها هنا، ومن يرغب في ذلك عليه الرجوع إلى انغوف (Angoff,1971).

الخطوة الأولى في المعادلة هي تحديد الرتب المئينية لتوزيع الدرجات على كلا الاختبارين (في التصميم ب تطبق صيغتي الاختبار على المجموعتين). ويبين جدول (20 –4) الرتب المئينية المقابلة للدرجات على صيغتي اختبار افتراضي، وبعدها تمثيل بياني لهذه الرتب مقابل الدرجات الخام على كلا الصيغتين، ومثل هذه المنحنيات موضح في الشكل (20 –2) لاختبار افتراضي مؤلف من (20) فقرة. وفي هذا الشكل يرسم خط يقطع كلا المنحنيين عند الرتبة المئينية نفسها، وبالطبع نرسم منحنى أملس يدوياً لتوزيع الدرجات التي حسبت لها الرتب المئينية، وهكذا يتم الحصول على الدرجات المتكافئة على المنحنى. ويوضح الشكل (20 –2) هذه الإجراءات للدرجة (12) على الاختبار (x) والدرجة المقابلة لها على الاختبار (y). ويمثل الجدول (20- 5) الرتب المئينية، ودرجات (x) ودرجات (y*) المناظرة لها، والمحسوبة باستخدام المنحنى في الشكل (20-2).

جدول (20-4): الرتب المئنية على صيغتي اختبار.

الدرجة	X	y
0	1	1
1	3	3
2	5	5
3	8	6
4	14	10
5	22	15
6	26	20
7	32	25
8	40	29
9	48	35
10	57	43
11	66	53
12	72	61
13	76	66
14	82	75
15	88	82
16	91	87
17	95	92
18	97	95
19	98	98
20	99	99



شكل (20-2): تمثيل بياني للرتب المئينية المتكافئة لصيغتي اختبار. افتراضيين ويتألف كل منهما من (20) فقرة.

جدول (20-5): الدرجات المئينية المتكافئة

Y*	X	الرتبة المئينية
0.0	0	1
1.0	1	3
2.0	2	5
2.3	3	6
3.3	4	10
4.2	5	15
4.7	6	20
5.6	7	25
6.4	8	29
7.3	9	35
8.3	10	43
9.4	11	53
10.1	12	61
11.0	13	66
12.8	14	75
14.0	15	82
15.2	16	87
16.2	17	92
16.9	18	95
19.0	19	9
20.0	20	99

وبعد تحديد كل من x، *y تمثل بيانياً وبمنحنى أملس عبر الدرجات، ويمكن رسم منحنى مثل المنحنى في الشكل (1-20) يدوياً أو بطريقة تحليلية باستخدام الطريقة التي عرضها كولين (Kolen, 1984)، وبعدها تقرأ الدرجات المتكافئة عن المنحنى، وفي مثالنا تكون درجات *y تقريباً كما في جدول (20-5)، لهذا فلن نعرض جدول آخر للدرجات المتكافئة.

ولبيان جدول الدرجات المتكافئة، يمكننا تحويل درجات *y إلى درجات Y كما هو موضح في الجدول (20-6)، وكمثال لتفسير هذا الجدول: حصل فرد على الدرجة (13) في اختبارY تكون درجته على اختبار x = 14، والرتبة المئينية لكل منهما تساوي 75.

جدول (20-6): الرتب المئينية المتكافئة على صيغتي اختبار افتراضيين يتألف كل منهما من (20) فقرة.

الرتبة المئينية	X	Y^a
1	0	0
3	1	1
5	2	2
6	3	3
10	4	4
15	5	5
20	6	6
25	7	7
29	8	8
35	9	9
43	10	10
53	11	11
61	12	12
66	13	13
75	14	14
82	15	15
87	16	16
92	17	17
95	18	18
98	19	19
99	20	20

a: تم الحصول على الدرجات بتحويل درجات *y في جدول (20-5).

الاختيار عبر طرائق المعادلة:

يوجد ثلاثة محكات لاختيار طريقة المعادلة: الافتراضات ودرجة التقيد بها، والتطبيق، والدقة. وهذه المحكات تطبق على كل من اختيار المعادلة الخطية أو معادلة الرتب المئينية لتصميم معين وللاختيار عبرالتصميمات.

أولاً: افترض أن الاختيار بين المعادلة الخطية والرتب المئينية: المناقشة الآتية تطبق على التصميمات الثلاثة: تتضمن معادلة الرتب المئينية عدد من الافتراضات أقل من افتراضات المعادلة الخطية، إذ أن المعادلة الخطية تفترض أن الفرق بين توزيعي X و Y في المتوسط والتباين فقط. واستخدام عدد أقل من الافتراضات يعد أمر مهم. ومع ذلك فإن الرتب المئينية تتطلب عددا من الرسوم البيانية، وإجراؤها يكون أكثر تعقيداً من المعادلة الخطية، لذا فإنها عملية لدرجة أقل.

وثانياً: يكون خطأ المعادلة النظري أكبر بكثير من المعادلة الخطية (lord,1982). لهذا فإنه يفضل استخدام المعادلة الخطية فيما لو كان توزيع الدرجات الزائية لكل من x و y لا يختلف كثيراً.

ويعد الاختيار العملي عبر التصميمات المحك الأساسي، فالتصميمات (أ و ب) تتطلب تداخل بسيط أو معدوم بين الاختبارين، وهذا مستحيل من الناحية العملية، لذلك فإن التصميم (ج) يفضل من الناحية العملية. ويظهر موقف مثل هذا عند استخدام صيغ اختبارية عديدة لاختبار ما خلال العام الدراسي، كذلك فإن قضية عدم تسرب الأسئلة تتطلب إجراء دراسة معادلة الصيغ الاختبارية بمعادلة مستقلة.

كذلك فإن التصميم (ج) يعد عملياً بدرجة أكبر في مواقف كهذه، فطريقة المعادلة تعتمد على الافتراضات الخاصة بالتصميمات (أ و ب) إضافة إلى افتراضات أخرى خاصة. وحذّر انغوف (Angoff,1971) من عدم التطابق لكلا المعادلتين الخطية والمئينية فيما لو كان توزيع درجات الاختبار المشترك لكلا المجموعتين الفرعيتين مختلف بشكل واضح. وقد وصفت طريقة بديلة لمثل هذا الموقف، إضافة إلى إمكانية تطبيق الطرائق التي تعتمد نظرية السمة الكامنة، ولكن استخدام منهجية انغوف أو تلك التي تعتمد نظرية السمة الكامنة تتطلب افتراضات أخرى بالإضافة إلى الافتراضات التي تتطلبها تصميمات (أو ب). وستطرح منهجية نظرية السمة الكامنة في المعادلة هذا الفصل لاحقاً.

وعندما يكون بناء اختبارين بتداخل زمني قليل ممكناً يمكننا استخدام التصميمات الثلاثة في المعادلة، وفي موقف كهذا يمكن توزيع المجموعة عشوائياً عند استخدام التصميم (ج)، وهذا افتراض عُرض مسبقاً. وعند الاختيار عبر التصميمات الثلاثة يجب أن يؤخذ بعين

الاعتبار كلاً من قضيتي الدقة والتطبيق ليكون الاختيار منطقياً. ولتوضيح ذلك وازن بين التصميمين (أ و ب). ففي التصميم (أ) يتقدم المفحوص لإحدى الصيغتين، في حين يتقدم لكليهما في التصميم (ب) لذا فإن التصميم (أ) يتطلب وقت أقل لكلا مفحوص، ويعد هذا جزءاً من القضية فقط.والخطأ المعياري لكل التصميمين على النحو الآتي:

$$\frac{2\sigma_y^2}{N} \ (Z^2_x + 2) \ = SEy^* \qquad (5\text{-}20)...............$$

$$\frac{\sigma^2 y\ (1-\hat{\rho}_{XY})}{N} \ \ [Z^2_x\ (1+\hat{\rho}_{xy}) + 2] = SEy^* \qquad (6\text{-}20)...............$$

وتمثلN في هاتين المعادلتين إلى عدد المفحوصين الكلي و Z_x إلى الدرجات الزائية لدرجة x المحولة على y*. وبالمقارنة بين المعادلتين (20-5) و (20-6) يبدو أنه إذا كانت N هي نفسها في كلا التصميمين فإن الخطأ المعياري للتصميم ب أقل، كذلك فإنه يكون أقل عندما تكون $\hat{\rho}_{XY}$ أكبر. والنسبة بين عدد المفحوصين تتطلب الحصول على دقة متساوية.

$$\frac{2\ (Z^2x + 2)}{(1-\hat{\rho}_{XY})\ [Z^2x\ (1+ \hat{\rho}_{XY}) + 2]} = \frac{NA}{NB} \qquad (7\text{-}20)...............$$

على سبيل المثال: افترض أن $\hat{\rho}_{xy} = 0.8$، وللحصول على الدقة نفسها لدرجة x التي قيمة Z_x لها = صفر، فهذا يتطلب أن تكون النسبة:

$$\frac{10}{1} = \frac{(2+ 20)\ 2}{[\ 2 + (0.8 + 1)\ 20\ 6(0.8 - 1)} = \frac{NA}{NB}$$

وهذا يعني أنه يتطلب منا عشرة أضعاف العدد من المفحوصين للتصميم (أ)، لذا فإن التصميم (أ) يكون عملياً فيما لو توافر عدد كبير من المفحوصين والاختبارين طويلين لكل مجموعة مفحوصين، بينما يكون التصميم (ب) عملياً إذا كان عدد المفحوصين أقل والاختبارات قصيرة لدرجة كافية لتطبيقها على المفحوصين في كلا المجوعتين.

والمقارنة الأخرى عبر التصاميم على النحو الآتي: كلما كان الارتباط أكبر بين الاختبار المشترك وكلا الصيغتين كلما كانت النتائج أكثر دقة في التصميم (ج) من تلك الناتجة عن التصميم (أ). ومع ذلك فإن التصميم (ج) يتطلب وقت أطول للتطبيق لكل مفحوص. واقل عدد من الفقرات المشتركة المثبت تجريبياً هو (20) فقرة أو (20%) من عدد فقرات أي من

الاختبارين (Angoff,1971). لذا فإن الاختبار المشترك يحتاج إلى وقت اختباري إضافي. وصيغة حساب الخطأ المعياري للتصميم (جـ) هي نفسها المستخدمة للتصميم (ب)، لذا فإن استخدام الاختبار المشترك يزيد من درجة الدقة. وعند المقارنة بين التصميمين (ب و جـ) فإن التصميم (ب) يتطلب وقت اختباري أطول من التصميم (جـ) ولكل مفحوص. وعموماً فعندما يكون الارتباط بين الاختبارين أكبر من الارتباط بين الاختبار المشترك مع كلا الصيغتين فإن المعادلة باستخدام التصميم (ب) تكون أكثر دقة من استخدام التصميم ج.

المعادلة باستخدام نظرية السمة الكامنة:

كما أشرنا مسبقاً فإنه يمكن استخدام نظرية السمة الكامنة في معادلة الدرجات وباستخدام التصميمات الثلاث. وتكون أكثر فائدة على الأغلب مع لتصميم (جـ) عندما لا يكون توزيع الدرجات عشوائي. وهنا فإنه ليس من الضروري توافر الافتراضات اللازمة لكل من المعادلة الخطية والمئينية. وسنفترض في هذا الجزء أن المعادلة ستجري باستخدام التصميم (جـ) ودون التعيين العشوائي لكلا المجموعتين. كذلك فإنه ليس من الضروري استخدام نظرية السمة الكامنة عندما يكون التوزيع عشوائي لتصميم (جـ) وللتصميمين (أ و ب)، ففي هذه الحالات تعد المعادلة الخطية والمئينية ملائمة.

ويتوافر حالياً العديد من نماذج السمة الكامنة، والنموذج الأكثر شيوعاً هو نموذج راش اللوغاريتمي أحادي المعلم، والنموذج اللوغاريتمي ثلاثي المعلم. ويمكن إجراء المعادلة باستخدام أي منهما، بالإضافة لاستخدام النموذجين اللوغاريتمي ثنائي المعلم واللوغاريتمي الطبيعي. ولتسهيل العرض قد اعتمدنا النموذج اللوغارتيمي أحادي البُعد (المعلم) ومع هذا التبسيط فإن العرض ليس تاماً. ولتفاصيل أكثر يمكن العودة إلى (Wright&stone, 1979) و (Lord,1980) الذي يبحث موضوع المعادلة باستخدام النموذج اللوغارتيمي ثلاثي المعلم.

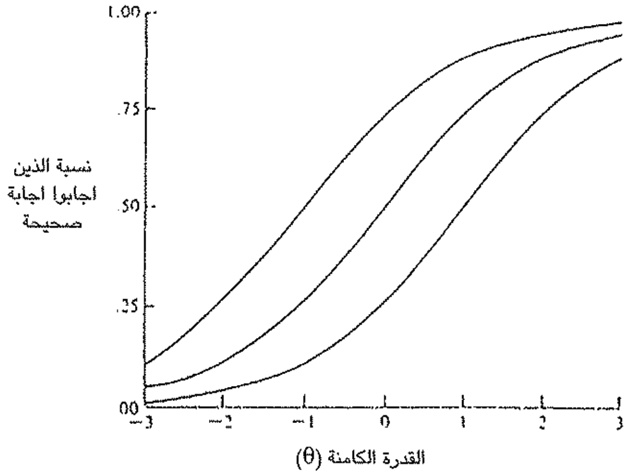
مدخل مختصر لنظرية السمة الكامنة:

للقارئ الذي لم يطلع على الفصل الخامس عشر نقدم عرضاً موجزاً لنظرية السمة الكامنة هنا. تفترض هذه النظرية أن سمة كامنة واحدة تفسر العلاقة بين الأداء (الاستجابات) على الفقرات جميعها. وهذا النموذج يبين العلاقة بين الدرجات على الفقرة ودرجة توافر السمة الكامنة. والنموذج اللوغاريتمي أحادي البُعد أو كما يُعرف بنموذج راش ينص على أن نسبة

المفحوصين الذين أجابوا إجابة صحيحة على عدد من الفقرات g هي [$p_g(\theta)$] يرتبط بالدرجة على السمة الكامنة كما تمثله المعادلة:

$$p_g(\theta) = \frac{e^{(\theta - bg)}}{1 + e^{(\theta - bg)}} \quad ....(9\text{-}20)$$

ويبين الشكل(20-3) تمثيلاً بيانياً للعلاقة بين سمة كامنة افتراضية لثلاثة فقرات تختلف في قيمة p_g، ويطلق على هذه المنحنيات اسم المنحنى المميز للفقرة (ICC)، ويمثل المعامل b_g معامل صعوبة الفقرة، مع ملاحظة أنه ليس معامل الصعوبة نفسه في النظرية التقليدية. وسيجيب 50% من المفحوصين الذين حصلوا على درجة سمة كامنة bg على الفقرات g_{th} إجابة صحيحة. فمثلاً إذا كانت b_g =1 فإن 50% من المفحوصين الذين درجة سمتهم الكامنة =1 سيجيبون على الفقرات g_{th} إجابة صحيح. ويتبين من شكل (20-3) أن قيمة b_g للمنحنى المميز للفقرة الأولى هو الأقل، إذ تصل قيمته إلى 5.0 قبل المنحنيين الآخرين. لاحظ وبالعودة إلى المعادلة (20-10) أنه إذا كان هنالك فقرتين لهما معامل الصعوبة نفسه فإن المنحنى المميز لهما يجب أن يكون نفسه، وهذا صحيح لأن المنحنى المميز للفقرة يعتمد فقط على θ وعلى درجة السمة الكامنة b_g. ومن هنا فإن الفقرات تختلف فقط في درجة صعوبتها، ويفترض أنها جميعاً متساوية في القوة التمييزية، وباختصار نقول أن هذا يكافئ الافتراض بأن ارتباط الفقرة والدرجة الحقيقية هو نفسه للفقرات جميعها.



شكل (20-3): المنحنى المميز لثلاث فقرات

تدريج درجات السمة الكامنة: Scaling latent Trait Scores

في العديد من القياسات الفيزيائية مثل الطول والوزن تمثل نقطة الصفر في التدريج غياب السمة المقيسة، ونتيجة لذلك فإن إعطاء درجة الصفر للشيء لا يعد تمثيلاً مناسباً. فالدرجة صفر تعطى فقط للأشياء التي لا تمتلك شيئاً من الخاصية المقيسة. وفي نظرية السمة الكامنة لا يمثل الصفر غياب السمة، وعلى هذا فإن الصفر يمكن استخدامه مع أي مجموعة من المفحوصين المتجانسين في السمة الكامنة، وعندما تشير إلى أن درجة المفحوصين تساوي صفر فليس ضرورياً تحديد الدرجات الاخرى ويجب اعطاء المفحوصين الذين يمتلكون سمة كامنة بدرجة أكبر درجات أكبر من صفر وهكذا.

وطريقة أخرى لتحديد نقطة الصفر هي أن إعطاء أي تدريج θ لدرجات السمة الكامنة: يمكننا من تحويل θ إلى *θ، وتحويل $\theta^* = \theta + m$ وتشير قيم m المختلفة إلى نقطة الصفر لمجموعات المفحوصين المختلفة. لاحظ أن القيمة العددية لـ bg هي قيمة على تدريج السمة الكامنة، فعندما يتم تحويل θ على*θ فإن bg تحول إلى b_g^* وتساوي $b_g + m$.

وتوجد طريقة شائعة لتحديد نقطة صفر مناسبة، وهي تعيين تدريج بحيث يكون معدل درجة السمة الكامنة لمجموعة أو مجتمع معين تساوي صفر. دعنا نرى مطابقة هذه الطريقة في تحديد نقطة الصفر، افترض أنه تم تطبيق مجموعة فقرات على مجموعتين من المفحوصين، وحيث أن نقطة الصفر مناسبة فهي تعطى لكل مجموعة بصورة منفصلة. ويتساوق هذا مع اختيار التدريج θ_1 للمجموعة الأولى، وبهذا يكون متوسط السمة عندها = صفر، وباختيار θ_2 على أنها تدريج للمجموعة الثانية فإن متوسط السمة الكامنة على التدريج الجديد = صفر. وحيث أن كلا التدريجين يقيس السمة الكامنة نفسها فإن التدريجين يرتبطان بالعلاقة $\theta_1 = \theta_2 + m$، حيث أن θ_1 هي تدريج المجموعة الأولى و θ_2 هي تدريج المجموعة الثانية. وتعني هذه المعادلة أنه لتحويل الدرجات من تدريج المجموعة الثانية إلى درجة على تدريج المجموعة الأولى فمن الضروري إضافة m إلى الدرجة المحوّلة.

افترض أنه من الممكن تحديد مفحوصين اثنين واحد من كل مجموعة ويمتلكون الكمية نفسها من السمة الكامنة أو القدرة الكامنة، وبما أن θ_1 هي درجة المفحوص في المجموعة الأولى و θ_2 هي درجة المفحوص من المجموعة الثانية. وكل منهما يمثل الكمية نفسها من القدرة ولكنها ممثلتان على تدريجين مختلفين، لذا فإن $b_{g1} - b_{g2} = m$. وبالحصول على قيمة m فإنه يمكن تحويل أي درجة على تدريج θ_2 إلى درجة على تدريج θ_1 باستخدام المعادلة:

$$m + \theta_2 = \theta_1$$

وعند استخدام التصميم (جـ) فإن مجموعة فقرات (الاختبار المشترك) تعطى لكلا

المجموعتين، ولأي من هذه الفقرات فإن $b_{g1} - b_{g2} = m$. ويستمر استخدام m في تحويل درجة من تدريج المجموعة الثانية إلى درجة على تدريج المجموعة الأولى. ويمكن أيضاً تقدير قيم كل من b_{g1} و b_{g2} عملياً.وكنتيجة فإن $(b_{g1}-b_{g2})$ ليست القيمة نفسها لكل فقرة في الاختبار المشترك، ويؤخذ عادة معدل هذه الفروقات للحصول على تقدير لقيمة m. وستعرض تفصيلات أكثر تحت العنوان التالي "طبيعة الاختبار المشترك".

طبيعة الاختبار المشترك

يتطلب التصميم (جـ) استخدام اختبار مشترك لصيغ الاختبارات المراد معادلتها. وفي حالة استخدام نظرية السمة الكامنة مع البيانات المجمعة للتصميم (ج) فمن الضروري أن تقيس فقرات الصيغتين وكذلك الاختبار المشترك السمة نفسها، في حين أن هذا الشرط ليس ضرورياً في المعادلة الخطية والمعادلة المئينية للبيانات المستخدمة مع التصميم (ج).

ولأن الفقرات جميعها تطبق على مجموعة بما فيها فقرات الاختبار المشترك وتقيس السمة الكامنة نفسها فمن الممكن استخدام استجابات الفقرات جميعها للحصول على درجة المفحوص في الاختبار، وهذا ما سنعتمده في وصف معادلة الدرجات باستخدام نظرية السمة الكامنة. ومع ذلك فإن بعض الشركات التجارية تستخدم الاختبار المشترك لأغراض المعادلة فقط، أي أن درجات الاختبار المشترك لا تؤثر على درجة المفحوص.

خطوات المعادلة:

أن الطريقة الأبسط لوصف طريقة المعادلة هي عرض الخطوات ثم تفسيرها:

1/ قدِّر قيم b_g للفقرات جميعها بما فيها الاختبار المشترك باستخدام بيانات المجموعة الأولى. هذه الخطوة تم تعديلها بناءً على افتراض أن الفقرات جميعها بما فيها فقرات الاختبار المشترك تقيس السمة الكامنة نفسها، وعند إجراء التقدير يتم تدريج هذه التقديرات بحيث يكون متوسط درجة السمة الكامنة = صفر. ويبين جدول (20-7) تقديرات قيم b_g لخمس عشر فقرة أخذت للمجموعة الأولى والفقرات الخمس الأخيرة هي فقرات الاختبار المشترك (ويستخدم برنامج بيكال BICAL الذي أعده رايت وزملاؤه (Wright&Mead& Bell,1979) في حساب هذه التقديرات، حيث يتم تحليل البيانات للمجموعتين بصورة منفصلة. وعلاوة على اختيار متوسط درجة السمة الكامنة مساوياً لصفر فإن برنامج بيكال يختار التدريج بحيث يضع متوسط درجات السمة الكامنة المعينة لكل درجة مصححة تساوي صفر. ولا تغير هذه الطريقة بأي حال من الأحوال بشكل هام التعديل المستخدم لتقدير m عن طريق الفروقات بين قيم b_g للاختبار المشترك).

2/ باستخدام بيانات المجموعة الثانية قدِّر قيم الفقرات جميعها بما فيها فقرات الاختبار المشترك. ويتم تدريج الفقرات هنا باعتبار متوسط درجة السمة الكامنة = صفر، ويبين جدول (20-7) قيم b_g لخمس عشر فقرة الخمس الأولى منها هي فقرات الاختبار المشترك.

3/ احسب $b_{g1} - b_{g2}$ لفقرات الاختبار المشترك الخمس وبعدها احسب قيمة m بحساب معدل هذه الفروقات. وفي الجدول (20-7) تبيان لقيم $b_{g1} - b_{g2}$ لفقرات الاختبار المشترك جميعها في العمود الثالث، ومتوسط هذه القيم $m = 0.60$.

4/ قدر درجة السمة الكامنة المقابلة لكل درجة ولكل مجموعة، ويبين جدول (20 – 8) هذه التقديرات، وهذا أيضاً باستخدام برنامج بيكال الذي يحسب صعوبة الفقرة ويحوّل الدرجات الخام إلى درجات السمة الكامنة لكلا المجموعتين. وعلى القارئ أن ينتبه إلى أن درجات السمة الكامنة لكلا المجموعتين تعتمد على تدريج يختلف عن الأخر.

5/ أضف قيمة m إلى درجات السمة الكامنة للمجموعة الثانية لتحويلها إلى درجات على تدريج المجموعة الأولى، فمثلاً الدرجة الخام (1) في المجموعة الثانية لها درجة سمة كامنة (– 2.72) كما يبين هذا جدول (20 –8). ولتحو يل هذه الدرجة إلى التدريج المستخدم مع المجموعة الأولى أضف (0.60) إلى (–2.72) ونحصل على (-2,12). ويبين الجدول (20-9) الدرجات المحوّلة جميعها إضافة إلى درجات المجموعة الاولى الاصلية، وكلاً من درجات السمة الكامنة على التدريج نفسه، معادلة طبعاً.

جدول (20 -7) تقدير صعوبة الفقرات للمجموعة الأولى والثانية :

الفقرة	تقدير الصعوبة		الفرق بين درجات فقرات الاختبار المشترك
	مجموعة1 (صيغة X)	مجموعة2 (صيغة Y)	
1	1.867-		
2	0.727-		
3	0.603		
4	0.061		
5	0.016		
6	1.328-		
7	0.223-		
8	0.321		
9	0.223-		
10	0.483		
11	0.030-	1.85-	0.155
12	0.873	0.347	0.526
13	0.797	0.159	0.638
14	0.797	0.45	0.752
15	0.444	0.503-	0.947
16		0.302-	
17		0.342-	
18		0.159	
19		0.714-	
20		0.069-	
21		0.158-	
22		0.928	
23		0.159	
24		0.691	
25		0.185-	

جدول (20-8): جدول تحويل الدرجات الخام الى درجات السمة الكامنة

الدرجة الخام	درجات السمة الكامنة مجموعة 1 (صيغة x)	مجموعة 2 (صيغة y)
1	2.91-	2.72-
2	2.07-	1.93-
3	1.53-	1.43-
4	1.12-	1.04-
5	0.76-	0.72-
6	0.45-	0.42-
7	0.15-	0.14-
8	0.15	0.14
9	0.45	0.42
10	0.76	0.72
11	1.12	1.04
12	1.53	1.43
13	2.07	1.93
14	2.91	2.72

جدول (20 -9) جدول تحويل الدرجات الخام الى درجات السمة الكامنة المعادلة (أ)

الدرجة الخام	درجات السمة الكامنة مجموعة1	مجموعة 2
1	2.91-	2.12-
2	2.07-	1.33-
3	1.53-	0.83-
4	1.12-	0.43-
5	0.76-	0.12-
6	0.45-	0.18
7	0.15-	0.46
8	0.15	0.74
9	0.45	1.02
10	0.76	1.32
11	1.12	1.64
12	1.53	2.03
13	2.07	2.53
14	2.91	3.32

(أ) تم التعبير عن الدرجات جميعها بتدريج المجموعة الاولى

معادلة الدرجات الحقيقية :

حدد لورد(Lord , 1980) طريقة للمعادلة باستخدام نظرية السمة الكامنة تعرف بـ " معادلة الدرجات الحقيقية). وسنفسر في هذا الجزء معادلة الدرجات الحقيقية باستخدام المثال المطروح في الجزء السابق، ويلزم لمعادلة الدرجات الحقيقية تحديد الدرجات الحقيقية المتكافئة على صيغتي الاختبار. والدالة التي تربط الدرجات الحقيقية تستخدم فيما بعد لمعادلة الدرجات الملاحظة. وتبين المعادلة (20–10) العلاقة بين درجات السمة الكامنة والدرجات الحقيقية:

$$\sum_{g=1}^{G} P_g(\theta) = T \qquad (10\text{-}20)\ldots\ldots\ldots\ldots\ldots$$

وتشير هذه المعادلة الي انه ولأي قيمة θ يمكن تحديد الدرجة الحقيقية المقابلة لها، وذلك بحساب $P_g(\theta)$ لعدد من الفقرات = G في الاختبار، وبجمع جميع الحدود، وعلى سبيل المثال الدرجة الحقيقية T_x المقابلة لـ $\theta = 1.5$ هي:

$$\sum_{g=1}^{15} P_g(1.5) = T_x \qquad (11\text{-}20)\ldots\ldots\ldots\ldots\ldots$$

وتذكر أن:

$$\frac{e^{(1.5 - bg)}}{1 + e^{(1.5 - bg)}} = P_g(1.5) \qquad (12\text{-}20)\ldots\ldots\ldots\ldots$$

وبتعويض تقديرات bg من جدول (20-7) نحصل على:

$$11.9 = \frac{e^{(1.5 - b1)}}{1 + e^{(1.5 - b1)}} + \ldots\ldots\ldots\ldots\ldots + \frac{e^{(1.5 - b15)}}{1 + e^{(1.5 - b15)}} = Tx$$

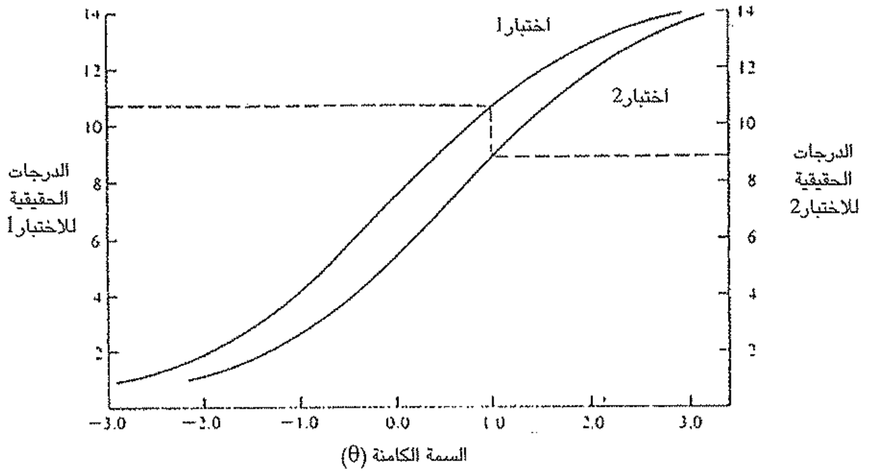
$$11.9 = \frac{e^{(1.5 - 1.867)}}{1 + e^{(1.5-1.867)}} + \ldots\ldots\ldots\ldots\ldots + \frac{e^{(1.5-0.444)}}{1 + e^{(1.5-0.444)}} =$$

ومنها فإن درجة السمة الكامنة على الصيغة x (1.5) يقابلها درجة حقيقية قيمتها 11.9.

ويبين الشكل (20-4) منحنى العلاقة بين T_x و θ وكذلك بين T_yو θ، وكل منهما يطلق عليه المنحنى المميز للاختبار. ويمكن اعتبار درجتين حقيقيتين (احداهما على الصيغة x والاخرى على الصيغة y) متكافئتين اذا كانتا مقابلتين لدرجة السمة الكامنة نفسها ويبين الخط

المتقطع في الشكل (20-4) الدرجات الحقيقية المقابلة لـ $\theta = 1.5$، ويبين الشكل (20-5) تمثيلاً بيانياً للدرجات الحقيقية المتكافئه على كلا الاختبارين.

واذا كان ممكنا قياس مفحوص على الصيغة x والحصول على درجته الحقيقية فيمكننا باستخدام منحنى الدرجة الحقيقية المكافئة تحديد درجته الحقيقية على الصيغة y ، مع اننا لم نجري له قياساً على الصيغة y ، وهكذا فإن منحنى الدرجات الحقيقية المتكافئة يمثل معادلة للدرجات الحقيقية. وتكمن المشكلة الحقيقية بعدم امكانية الحصول على الدرجة الحقيقية على الاختبار x، واقترح لورد استخدام منحنى الدرجات الحقيقية المتكافئة بتعويض الدرجات الخام بدلاً من الدرجات الحقيقية (واقتراح اخر بديل من قبل لورد هو استخدام نظرية السمة الكامنة في تقدير الدرجة الحقيقية وبعد ذلك تعويضها بالدرجات الحقيقية الفعلية) وعندئذ تستخدم الدرجات الحقيقية المتكافئة على انها دالة $y^* = f(x)$، والتي بوساطتها تحول الدرجة x الى الدرجة الكافئة على تدريج الصيغة y.

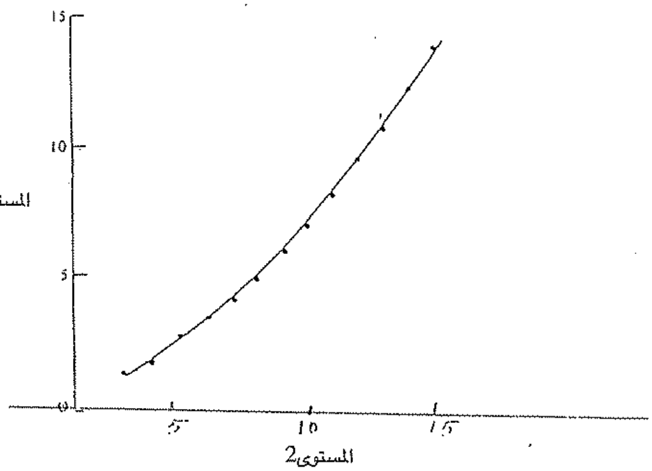


شكل(20-4): العلاقة بين الدرجات السمة الكامنة والدرجات الحقيقية على مستويي الاختبار

المعادلة العمودية:

وتستخدم لمعادلة درجات اختبارين او اكثر يعتقد انهما يقيسان السمة الكامنة نفسها مصممة بحيث تكون مختلفة في درجة الصعوبة وقد اشرنا في بداية الفصل الى احد اسباب استخدام المعادلة العمودية واللازم لتوحيد.اختبارات مستويات مختلفة. وقد أشرنا الى انه في

حالة بطارية الاختبارات المتضمنة مستويات عديدة تناسب صفوف مختلفة قد يكون مفيداً لبعض الطلبة اختبارهم باختبارات تناسب صفوف تختلف في مستواها عن صفه الفعلي، فعلى سبيل المثال الطالب المتفوق جداً في نهاية الصف الثالث يمكن اختباره باختبار يلائم بداية الصف الخامس، وكذلك فللطالب الذي لديه صعوبات تعلم في نهاية الصف الثالث يمكن اختباره باختبار يلائم نهاية الصف الثاني. وفي كلا الحالتين فمن الضروري معادلة درجات الاختبارات من المستويين في حالة اعطاء الطالب اختبار خارج مستواه، وذلك عند مقارنته مع اقرانه الذين تقدموا لاختبار بمستوى الصف الفعلي نفسه.



شكل (20-5): تمثيل بياني للدرجات الحقيقية المتكافئة على مستويي الاختبار

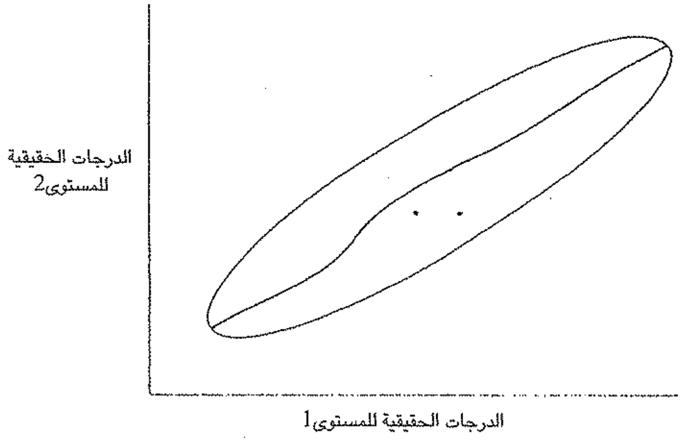
وسبب اخر لمعادلة المستويات المختلفة مشتق من الحاجة الى مخطط نمو الطفل عبر سنوات الدراسة والمرغوب فيه هو اشتقاق تدريج يبين درجات مستويات بطارية الاختبارات التحصيلية جميعها، وهذا يتطلب معادلة المستويات، واذا ما تم هذا التدريج عبر انتقال الطفل من صف لاخر وتقدم لاختبارات بمستويات اكثر تقدماً، فالدرجة على الاختبار الحالي يمكن مقارنتها بدرجته على الاختبارات ذات المستوى الأدنى، والتدريج المناسب لمثل هذه الاختبارات المعيارية يتطلب المعادلة العمودية.

ويمكن اجراء المعادلة العمودية باستخدام اي من التصميمات (أ أو ب أو جـ) وخطياً او الرتب المئينية. ففي السبعينات استخدم التصميم أ مع معادلة الرتب المئينية في بناء تدريج العديد من بطاريات الاختبارات التحصيلية. والاكثر حداثة هو استخدام معدي الاختبارات طرائق ونماذج السمة الكامنة في اجراء المعادلة، ومثال على ذلك هو التدريج المستخدم في

اختبار النسخة الاصلية المعدلة (1982) لاختبار ستانفورد التحصيلي، والذي اجريت معادلته باستخدام النموذج اللوغاريتمي احادي المعلم، وكذلك معادلة تدريج نسخة (1978) لاختبار المهارات الاساسية الشامل وباستخدام النموذج اللوغاريتمي ثلاثي المعلم

وهناك قضية يجب ان تؤخذ بعين الاعتبار عند اجراء المعادلة العمودية هي احتمالية كون الاختبارات مختلفة اساساً في صعوبتها وفي السمة التي تقيسها رغم تشابه المحتوى. فمثلاً في احد الاختبارات التحصيلية المعيارية وفي أحد مستوياته صنف الطالب على انه في مستوى صف (1.5 الى 2.4)، واخر صنف في مستوى من (2.5 الى 3.4)، وتوفر الاختبارات الفرعية في الرياضيات لهذين المستويين درجات على التدريج نفسه في المعادلة العمودية، مع انه يتم قياس 57 هدف في احد المستويات و 22 هدف مشترك لكلا المستويين، ومع ذلك لا يعد هذا دليلاً على ان الاختبارين يقيسان سمات مختلفة، فهذا دليل مقترح فقط.

وما هي توابع كون المستويين يقيسان سمات مختلفة؟ يمكننا الاجابة عن السؤال فقط اذا تمكنّا تحديد ما نعنيه بقياس مستويين للسمة نفسها اجرائياً ولهذا الحد فإننا نقترح ان مستويين يقيسان السمة نفسها اذا كانت الدرجات الحقيقية على المستويين بينهما علاقة اقترانية كما هو مبين في الشكل (20-5). وفي هذه الحالة فإن معرفة درجة الطالب الحقيقية في احد المستويات يمكننا من تحديد درجته الحقيقية في المستوى الثاني وفي حالة عدم قياس المستويين للسمة نفسها فإن هنالك علاقة احصائية بين الدرجات الحقيقية كما هو موضح في التمثيل البياني في الشكل (20-6)، ويعبر هذا المنحنى عن معادلة الاقتران، والطريقة المستخدمة لاشتقاق معادلة الاقتران غير مهمة لاهداف المناقشة المبينة ادناه .



شكل (20-6) تمثيل بياني للدرجات الحقيقية على مستويي الاختبار

افترض ان هنالك مفحوصين اثنين لهماالدرجات الحقيقة نفسها في المستوى الثاني ولكنها مختلفة في المستوى الاول ومشار الى الدرجات الحقيقية لهذين المفحوصين بنجمة في الشكل (6-20)، وستشير الى المفحوص الذي درجته الى اليسار بالاول، وحيث ان هذا المفحوص له درجة حقيقية اقل في المستوى الاول من المفحوص الثاني فإنه سيحصل على درجة أقل في اختبار المستوى الأول. وعند تحويل درجات المفحوصين الاثنين الى درجات المستوى الثاني باستخدام معادلة الاقتران، فان المفحوص الاول ستكون درجته اقل من درجة المفحوص الثاني. وهذا يعني وضع المفحوص الاول في مكان غير ملائم، ويمكن وصف هذا بطريقة مختلفة نوعاً ما، وذلك اذا اعتبرنا درجة المفحوص الحقيقية تضعه في نقطة أدنى من المعادلة الاقترانية . وإن تقدم هذا المفحوص الى صيغة الاختبار الاولى وحولت درجته الى تدريج صيغة الاختبار الثانية ، والعكس صحيح لاي مفحوص تكون درجته الحقيقية اعلى وتقع في موقع اعلى في المعادلة الاقترانية.

وعلى القارىء ان يميز ان المعادلة الافقية ستكون خاطئة فيما لو كانت الاختبارات المراد معادلتها لا تقيس السمة نفسها. وفي الاحوال المختلفة فإن الاختبارات المعادلة افقيا تكون مصممة للهدف نفسه وبمستوى الصعوبة نفسه ويبدو من الضروري اعتبار الاختبارين يقيسان السمة نفسها، وفي بعض الحالات تستخدم المعادلة الافقية لمعادلة اختبارين مختلفين، وهنا نطرح السؤال نفسه حول ما اذا كان الاختبارين يقيسان السمة نفسها

واضافة الى احتمالية كون مستويي الاختبار لا يقيسان السمة نفسها، فقد تظهر مشكلات اخرى عند استخدام طرائق السمة الكامنة في المعادلة العمودية، وسنركز هنا على المعادلة بوساطة النموذج اللوغاريتمي احادي المعلم، والذي يعتمد على مجموعة افتراضات اساسية، هي

1. تقيس فقرات الاختبار (او الاختبار الفرعي) سمة كامنة احادية البعد.

2. تختلف الفقرات في صعوباتها وليس في تمييزها.

3. لا يحاول المفحوصين تخمين الاجابة.

ولان هذا النموذج مبني على افتراضات قوية، فالسؤال الذي يطرح نفسه هو : هل تعد دالة المعادلة المطورة لهذا النموذج جيدة فيما لو اهملت هذه الافتراضات. وبالطبع يعد اهمال هذه الافتراضات احد مصادر المهمة في المعادلة الافقية ويبدو أنه أكثر تأثيراً في المعادلة العمودية. وعلى سبيل المثال افترض مشروع معادلة مستويين الاول يناسب الصف الثاني والاخر يناسب الصف الثالث، ويتقدم هنا طلبة الصف الثاني لاختبار المستوى الاول بالاضافة

الى الاختبار المشترك، في حين يتقدم طلبة الصف الثالث لاختبار المستوى الثاني بالاضافة الى الاختبار المشترك. وفي هذه الحالة يكون الاختبار المشترك اصعب على طلبة الصف الثاني، لهذا فان التخمين يؤثر بدرجة مختلفة على معامل الصعوبة للاختبار المشترك لكلا الصفين. ولان المعادلة تعتمد في الاساس على تقدير الفروق في معامل الصعوبة لكلا الصفين في الاختبار المشترك فان تأثير التخمين لا يكون بالقدر نفسه لطلبة الصفين في المعادلة

وهناك العديد من الدراسات، منها المباشر مثل holmes, & , hoover , 1982, slinde & linn, 1978, 1982)، ومنها غير المباشر مثل Forsyth, saisangian and Glimer , 1982 ; whitely and Dawis , 1981 ; wright 1974) والتي تختبر استخدام النموذج اللوغاريتمي احادي المعلم في المعادلة وكما هو متوقع فقد اقترح اصحاب هذه الدراسات اراء مختلفة تتعلق بقيمة النموذج اللوغاريتمي احادي المعلم في المعادلة العمودية، ويمكن اشتقاق الاختلافات في الاستنتاجات من اربعة مصادر على الاقل، هي :

1. استخدام الدراسات لاختبارات مختلفة .
2. استخدام الدراسات لتصميمات مختلفة .
3. استخدام الدراسات لقياسات مختلفة للدرجات المتعادلة المتشابهة .
4. لا يوجد محك واضح في الدراسات للحكم على ملائمة عملية المعادلة.

ولتوضيح هذه الاختلافات دعنا نراجع دراسات هولز (1982) وسلايند ولين (1979)، وتتألف الطرائق المستخدمة في هذه الدراسات من ستة خطوات، وعلى النحو الآتي:

1. بناء او استخدام اختبارين مختلفين في صعوباتهما، ومجموعة فقرات تمثل اختبار مشترك، والتي تعد جزءاً من كلا الاختبارين او انها تكون اختباراً مستقلاً.
2. تطبيق الفقرات على عينة مناسبة .
3. تحديد مجموعتي القدرة: العالية والمنخفضة، وصنفت هذه المجموعات بالاعتماد على محك اخر غير درجات الاختبار المستخدم في الخطوة الاولى (دراسة سلايندولين) ولكنه مرتبط بالسمة المقيسة بهذه الاختبارات.
4. ومن البيانات لفئة القدرة المنخفضة في المستوى الاولى ولفئة القدرة العالية في المستوى الثاني اجريت معادلة الدرجات باستخدام الطريقة الموصوفة في هذا الفصل تحت عنوان " المعادلة باستخدام نظرية السمة الكامنة "
5. استخدام الدرجات الخام للسمة الكامنة باستخدام جدول تحويل مشابه لجدول (20-9)
6. مقارنة درجات السمة الكامنة لنتائج الاختبارين.

واستنج هولز ان النموذج احادي المعلم لا يزود بمعاني كافية للمعادلة العمودية عبر مدى القدرة الكلي، في حين ان سلايند ولين كانا اكثر تفاؤلاً في استخدام هذا النموذج

احدى الفروق بين الدراستين كانت في استخدام هولز لعينة مناسبة من الطلبة إذ لم يستخدم جدول تحويل الدرجات الخام الى درجات السمة الكامنة. وقد كرر هولز الخطوة السادسة في عينة الصدق، في حين ان سلايند ولين لم يكن بامكانهما استخدام هذا النوع من الصدق وتشير نتائج هولز الى ان المعادلة كانت أقل كفاءة عبر عينة الصدق عنها للعينة الأصلية. وفرق اخر بين الدراستين هو ان الفرق بين المستويات في دراسة هولز كانت أقل وبفروق اكبر في الصعوبة من الاختبارات المستخدمة بدراسة سلايند ولين اضافة الى ان هولز دون الفروق بين المتوسطات المعيارية لتقديرات السمة الكامنة عند كل نقطة من النقاط الثمانية لتدريج القدرة لكلا المستويين وتم احتساب الفروق المعيارية لكل مفحوص باستخدام المعادلة :

$$D = \frac{\hat{\theta}_2^2 + \hat{\theta}_1^2}{\sqrt{\hat{\sigma}_{\theta_1}^2 + \hat{\sigma}_{\theta_2}^2}} \qquad (13\text{–}20)$$

حيث ترمز θ_1 و θ_2 الى تقديرات السمة الكامنة للمستوى الاولى والثاني على التوالي، وترمز $\sigma\theta_2^2$ و$\sigma\theta_2^2$ الى تقديرات الخطأ المعياري لتقدير $\hat{\theta}_1$و $\hat{\theta}_2$ (فان كانت المعادلة تعمل بشكل مناسب فان θ_1 يجب ان تساوي θ_2. وقد دوّن سلايند ولين الفرق بين المتوسطات لتقدير السمة الكامنة $(\theta_1 - \theta_2)$ لكلا الصفين، ولكنهما لم يدونا فروق المتوسطات عند النقاط المختلفة لتدريج السمة الكامنة واخيراً يبدو ان استنتاجات هولز لم تكن مبنية على اساس حجم ونمط الفروقات التي لاحظها دون محاولة جادة في اظهار كيفية تأثير قيمة هذه الفروقات على القرارات المبنية على الاختبار، في حين ان سلايند ولين حاولا الحكم على الفروقات بدلالة الخطأ المعياري للقياس وعدد نقاط الدرجات الخام المكافئة لهذه الفروقات، ولكنهما لم يحاولا ربط هذه الفروقات بالاستخدام الفعلي للاختبارات. ونقطة مهمة تتعلق بهذه المناقشة حتى لو لم يكن هنالك فوارق بين دراسات كل من هولز وسلايند ولين فمن غير المرغوب به الوصول الى قرارات غير متكافئة حول ملائمة استخدام النموذج احادي المعلم أو اي نموذج اخر للسمة الكامنة، فكل مشروع او دراسة مشابهة يجب ان تتجه نحو التحقق من ملائمة المعادلة في مشروع أو بحث معين.

المكافئة في المعادلة: Equity in Equaling

وكما لاحظنا فان مشكلة المعادلة تكمن في تحديد دالة المعادلة $f(x) = y^*$ والتي تستخدم في تحويل الدرجات x لما يكافئها على تدريج y (وبالقليل من الطرق المتبعة في نظرية السمة الكامنة لتحويل درجات x و y الى تدريج السمة الكامنة، وبقليل من التغيرات فإن المناقشة ادناه تنطبق على هذا النوع من المعادلة) ان هدف اشتقاق دالة المعادلة هو تحويل درجات المفحوصين على الاختيار x إلى درجات *y بحيث يمكن مقارنة درجات المفحوصين الذين تقدموا للاختبار y مع درجات المفحوصين على الاختبار x. والسؤال المطروح هنا هو ما المحك الملموس او المنطقي في مقارنة y مع *y.

لقد طرح لورد (**Lord,1980**) قضية محك المقارنة، واشار الى ان المحك يستخدم من اجل المكافئة او المساواة في المعادلة. ويوجد هنا محكان مهمان في هذه الحالة: الاول هو ان درجة المفحوص الحقيقية على y يجب ان تساوي درجته الحقيقية *y، والاخر ان توزيع درجات الخطأ للمفحوصين يجب ان يكون نفسه لكل من y و *y وان لم يتحقق اي من المحكين فإن فرصة المفحوص للحصول على درجة ما تعتمد على الاختبار الذي تقدم له، وهنا لا تقارن درجات y و *y ولا نعادل درجات x و y . وفي حالة توافر المحكين السابقين تعد كل من y و *y قياسات متوازية للسمة نفسها ولها التوزيع نفسه وعند الرجوع الى هذه المحكات يبرز السؤالين الآتيين:

1. هل يمكن لتطابق بيانات درجات الاختبار ان يحقق هذه المحكات؟ .
2. هل من الممكن اجراء اختبار للتحقق من تحقيق هذه المحكين؟

وقد وضح لورد بالنسبة للسؤال الاول انه لو كانت الاختبارات المراد معادلتها تقيس سمة احادية البعد لا يمكن ان تتوافق درجات y و *y مع المحكات الا اذا تم بناء كل من اختبارات x و y بحيث يكون لكل فقرة في الاختبار x منحنى السمة نفسه لكل فقرة في الاختبار y (ICC)، ومثل هذه الاختبارات ينتج عنها اختبارات متوازية، وبما انها متوازية اصلا فلا داعي لمعادلتها لانها متكافئة اصلا

هل نتائج لورد تعني انه من المستحيل ان تكون الدرجات المعادلة متكافئة؟

اولا: السمة الكامنة معرفة نظرياً فقط، لذا فان نظرية السمة الكامنة ليست صحيحة تماماً .

وثانيا: حتى لو كانت المضامين صحيحة تقريباً للاختبارات الطويلة فالاختلاف بين الصيغة والاخرى في الدرجات الحقيقية والاخطاء المعيارية الشخصية للقياس يمكن ان يكون ارتباطها غير منطقي

ثالثا: بالاعتماد على نظريات اخرى مثل النظرية التقليدية فمن الممكن ان تصبح الدرجات متكافئة بعد المعادلة للاختبارات غير المتوازية. وبالنسبة للسؤال الثاني فمن الممكن اختبار اتفاق كل من y *y مع المحك، فدرجات كل منهما لها معاملات ارتباط متساوية مع المتغيرات الاخرى جميعها، لذا فان فحص توافق اي من y و *y مع المحك يمكن اختباره بحساب معاملات ارتباطها مع متغيرات عديدة، كذلك فانه اذا عودلت ثلاثة اختبارات او اكثر فإنه يمكن استخدام التحليل العاملي المدوّر لاختبار امكانية تحقق الافتراضات.

الخلاصة:

تهدف المعادلة الى التوصل الى درجات متكافئة عبر ادوات مختلفة تقيس السمة نفسها. واذا عودلت درجات اختبارين، فان درجات المفحوصين على الاختبارين، يمكن مقارنتها حتى عندما يتقدم المفحوصين للاختبارات المختلفة ويوجد ثلاثة تصميمات للمعادلة، هي تصميم (أ، ب، جـ)، ففي التصميم (أ) يخضع المفحوص لاحد الاختبارات المراد معادلتها، ويتم تحديد الاختبار المطبق على المجموعة عشوائياً، بينما في التصميم (ب) فإن المفحوصين يتقدمون للاختبارات المراد معادلتها جميعا وترتيب تقديم الاختبارات للمفحوصين يتحدد عشوائياً وفي التصميم (ج) يتقدم المفحوص لاحد الاختبارات المراد معادلة درجاتها بالاضافة الى اختبار مشترك يتقدم اليه جميع المفحوصين، ويمكن اختيار المجموعة التي تتقدم لاحد الاختبارات عشوائياً، ولكنه ليس بالضرورة. كذلك فإن اختيار التصميم المناسب يخضع للسؤال حول مدى قابلية التصميم للتطبيق

وهنالك ثلاثة طرائق للمعادلة هي المعادلة الخطية ومعادلة الرتب المئينية المتساوية. ومعادلة السمة الكامنة. وكمبدأ يمكن استخدام اي منها مع اي من التصميمات الثلاث، وعمليا تستخدم الطرائق التي تعتمد نظرية السمة الكامنة على التصميم (جـ) في حالة عدم التحديد العشوائي للمجموعات المتقدمة للاختبارات المختلفة، ففي مثل هذه الحالة لا تكون افتراضات المعادلة الخطية ومعادلة الرتب المئينية منطقية، لذا فأن المعادلة باستخدام اي منهما لا تكون

دقيقة، ويكون البديل هو المعادلة بتطبيق نظرية السمة الكامنة مع التصميم (جـ) ودون التعيين العشوائي للمجموعات. وتستخدم كل من المعادلة الخطية والمئينية بشكل مناسب مع كلا التصميمين (أ و ب) ويعتمد الاختيار بين هاتين الطريقتين (أو من خلال الطرائق الثلاثة في حالة استخدام نظرية السمة الكامنة في المعادلة) على درجة الدقة في المعادلة. وهنالك اعتقاداً عاماً بأن توزيع الدرجات لاختبارات عدة متشابه تقريبا عدا المتوسط والانحراف المعياري، فإن المعادلة الخطية تكون اكثر دقة من معادلة الرتب المئينية المتساوية وعندما يكون الاختلاف في خصائص اخرى غير المتوسط والانحراف المعياري فان المعادلة باستخدام الرتب المئينية تكون اكثر دقة.

التمارين

1/ أعدباحث في التربية الخاصة صيغتين اختباريتين لقياس اتجاه الطلبة نحو الاطفال المنخرطين في التخصص الرئيسي. وقد اختار الباحث عشوائياً 10 معلمين لتطبق عليهم الصيغة أ و 10 معلمين اخرين لتطبق عليهم الصيغة ب وكانت البيانات على النحو الآتي:

درجات المعلمين على صيغتي اختبار الاتجاه

صيغة أ	صيغة ب
50	31
41	38
42	42
51	39
37	41
53	46
50	34
54	42
48	37
53	52

استخدم هذه البيانات في معادلة الصغتين (أ و ب).

2/ ارادت احدى شركات الاختبارات معادلة صيغتي اختبار تألفت كل منهما من (10) فقرات، وتم تطبيق كلا الصيغتين على المجموعة نفسها وفيما يأتي جدول تكراري للدرجات على كل صيغة اختبارية استخدم هذه البيانات لمعادلة درجات الصيغة (أ) لتصبح على تدريج درجات الصيغة ب.

الدرجة	الصيغة أ	الصيغة ب
0	0	13
1	2	19
2	17	54
3	28	36
4	39	39
5	50	40
6	65	49
7	56	27
8	27	17
9	11	12
10	4	5

3/ يمثل الجدول الاول معامل صعوبة الفقرة (معامل راش) على صيغتي اختبار تألفت كل منهما من (10) فقرات، ويتضمن كل منهما 5 فقرات مشتركة مع الصيغة الاخرى (الاختبار المشترك) وكانت الفقرات الخمس الاول للاختبار والخمس الثانية للاختبار المشترك. ويبين الجدول الثاني الدرجات الخام ودرجة السمة الكامنة المناظرة لها لكلا صيغتي الاختبار استخدم هذه البيانات في معادلة الدرجات من الصيغة أ الى الصيغة ب.

الجدول (1) : معاملات الصعوبة

صيغة أ	صيغة ب
0.739-	0.705-
0.739-	1.756
0.366-	0.051
1.799-	0.073
1.049-	0.705-

تابع جدول 1

0.188	1.197-
2.251	1.378
0.767	0.147-
0.767	0.147-
2.929	2.379

جدول 2 : الدرجات الخام و السمة الكامنة المقابلة :

الدرجة الخام	صيغة أ	صيغة ب
1	2.98-	2.89-
2	1.88-	1.82-
3	1.15-	1.12-
4	0.55-	0.53-
5	0.00	0.00
6	0.55	0.53
7	1.15	1.12
8	1.88	1.82
9	2.98	2.89

4/ طور باحث صيغتي اختبار نهائي لمساق ما لشعبتين مختلفتين تألف كل منهما من (50) فقرة (10) منها مشتركة بين الصيغتين و (40) فقرة خاصة بكل اختبار. وفيما يأتي درجات كل اختبار والفقرات المشتركة لكل شعبة. استخدم هذه البيانات في اشتقاق درجة لكل مفحوص بحيث تعكس الفقرات المشتركة والفقرات الخاصة بكل صيغة.

الدرجات على الاختبارات المستقلة والمشتركة على صيغتي اختبار نهائي

الجزء الأول		الجزء الثاني	
الفقرات المستقلة	الفقرات المشتركة	الفقرات المستقلة	الفقرات المشتركة
3	14	4	24
6	23	5	25
4	21	3	15
0	7	3	9
3	17	4	14
4	14	5	20
6	10	3	16
4	19	4	13
5	14	4	7
8	27	7	27

5/ خططت شركة اختبارات معادلة صيغتي اختبار باستخدام التصميم (ب) وبطريقة المعادلة الخطية للصيغتين، وسيتم تحويل الدرجات على الصيغة الأولى إلى تدريج درجات الصيغة الثانية، ومن البحوث السابقة فمن المنطق توقع أن الأرتباط المتبادل بين فقرات الصيغتين أن يكون 70. و وبتباين قدره 100 لأي من الصيغتين. وتأمل الشركة أن يكون الخطأ المعياري للمعادلة عند المتوسط للصيغة الأولى لا يزيد عن $\frac{1}{10}$ من الخطأ المعياري للقياس. فما حجم العينة التي يجب استخدامها لتحقيق هذا الهدف؟ وما عدد المفحوصين اللازم فيما لو استخدم التصميم (أ)؟

المراجع

قائمة المراجع

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

American Psychological Association (APA), American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME). (1974). *Standards for educational and psychological tests and manuals*. Washington, D.C.: American Psychological Association.

Anastasi, A. (1968). *Psychological testing* (3rd Ed.). New York: Macmillan.

Anderson, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123–140.

Anderson, L. W. (1981). *Assessing affective characteristics in the schools*. Boston: Allyn and Bacon.

Andrews, B. J., and Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement, 36*, 45–50.

Angoff, W. H. (1971). Norms, scales, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd Ed.). Washington, D.C.: American Council on Education.

Angoff, W. H. (1982). Uses of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.). *Handbook of methods for detecting item bias*. Baltimore: Johns Hopkins University Press.

Angoff, W. H., and Ford, S. F. (1973). Item race interaction on a test of scholastic aptitude. *Journal of Educational Measurement, 10*, 95–105.

Baglin, R. F. (1981). Does "nationally" normed really mean nationally? *Journal of Educational Measurement, 18*, 97–108.

Baker, F. B. (1981). A criticism of Scheuneman's item bias technique. *Journal of Educational Measurement, 18*, 59–62.

Behuniak, P., Jr. Archambault, F. X., and Gable, R. K. (1982). Angoff and Nedelsky standard setting procedures: Implications for the validity of proficiency test score interpretation. *Educational and Psychological Measurement, 42*, 247–252.

Berk, R. A. (1976). Determination of optimal cutting scores in criterion referenced measurement. *Journal of Experimental Education, 45*, 4–9.

Berk, R. A. (1978). The application of structural facet theory to achievement test construction. *Educational Research Quarterly, 3*, 62–72.

Berk, R. A. (1980a). A consumers' guide to criterion-referenced test reliability. *Journal of Educational Measurement, 17*, 323–350.

Berk, R. A. (1980b). Item analysis. In R. A. Berk (Ed.). *Criterion-referenced measurement: The state of the art*. Baltimore: Johns Hopkins University Press.

Beuchert, A. K., and Mendoza, J. L. (1979). A Monte Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement, 16*, 109–118.

Binet, A., and Simon, T. (1905–1908). The development of the Binet-Simon scale. In W. Dennis (Ed.). *Readings in the history of psychology*. New York: Appleton-Century-Crofts, 1948.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison Wesley.

Bliss, L. B. (1980). A test of Lord's assumption regarding examinee guessing behavior on multiple choice tests using elementary school children. *Journal of Educational Measurement, 17*, 147–153.

Bloom, Benjamin S. (Ed.). (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: McKay.

Bloom, B. S., Hastings, J. T., and Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bock, R. D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–460.

Bormuth, J. R. (1970). *On the theory of achievement test items*. Chicago: University of Chicago Press.

Boynton, P. L. (1933). *Intelligence, its manifestations and measurement*. New York: D. Appleton & Co.

Bray, J. H., and Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAS. *Review of Educational Research, 52*, 340–367.

Brennan, R. L. (1972). A generalized upper-lower item discrimination index. *Educational and Psychological Measurement, 32*, 289–303.

Brennan, R. L. (1980). Applications of generalizability theory. In R. A. Berk (Ed.), *Criterion-referenced measurement: The state of the art*. Baltimore: Johns Hopkins University Press.

Brennan, R. L. (1983). Elements of generalizability theory. Iowa City: The American College Testing Program.

Brennan, R. L., and Kane, M. T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement, 14*, 277–289.

Brennan, R. L., and Lockwood, R. E. (1980). A comparison of the Nedelsky and Angoff cutting score procedures using generalizability theory. *Applied Psychological Measurement, 4*, 219–240.

Brown, F. G. (1983). *Principles of educational and psychological testing*. New York: Holt, Rinehart, and Winston.

Brown, J. M., and Chang, G. (1982). The predictive validity of the Minnesota Reading Assessment in postsecondary vocational educational programs. *Educational and Psychological Measurement, 42*, 345–352.

Browne, M. W. (1975). Predictive validity of a linear regression equation. *British Journal of Mathematical and Statistical Psychology, 28*, 79–87.

Brownell, W. A. (1933). On the accuracy with which reliability may be measured by correlating test halves. *Journal of Experimental Education, 1*, 204–215.

Burke, C. J. (1963). Measurement scale and statistical models. In M. H. Marx (Ed.), *Theories in contempory psychology*. New York: Macmillan.

Burril, L. E. (1982). Comparative studies of item bias methods. In R. A. Berk (Ed.), *Handbook of methods for detecting item bias*. Baltimore: Johns Hopkins University Press.

Camilli, G. (1979). A critique of the chi-square method for assessing item bias. Unpublished paper, Laboratory of Educational Research, University of Colorado.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Campbell, D. T., and Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching*. Chicago: Rand McNally & Company.

Cardall, C., and Coffman, W. E. (1984). A method for comparing the performance of

different groups on the items in a test. (ETS Research Bulletin 64-61). Princeton, N.J.: Educational Testing Service.

Chissom, B. S., and Hoenes, R. L. (1976). A comparison of the ability of the D-48 test and the IPAT Culture Fair Intelligence Test to predict SRA achievement test scores for 8th and 9th grade students. *Educational and Psychological Measurement, 36*, 561–564.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.

Cleary, T. A., and Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement, 28*, 61–75.

Cole, N. S. (1973). Bias in selection. *Journal of Educational Measurement, 10*, 237–255.

Coombs, C. H. (1950a). The concepts of reliability and homogeneity. *Educational and Psychological Measurement, 10*, 43–56.

Coombs, C. H. (1950b). Psychological scaling without a unit of measurement. *Psychological Review, 57*, 145–158.

Cox, R. C. (1965). Item selection techniques and evaluation of instructional objectives. *Journal of Educational Measurement, 2*, 181–185.

Cox, R. C., and Vargas, J. S. (1966). A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Paper presented at the annual meeting of the National Council on Measurement in Education.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Cronbach, L. J. (1970). Review of *On the theory of achievement test items*. *Psychometrika, 35*, 509–511.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). *Educational Measurement* (2nd Ed.). Washington, D.C.: American Council on Education.

Cronbach, L. J. (1973). Disciplined inquiry. In H. S. Broudy (Ed.). *Philosophy of educational research*. New York: John Wiley.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. *American Psychologist, 30*, 1–14.

Cronbach, L. J. (1976). Equity in selection: When psychometrics and political philosophy meet. *Journal of Educational Measurement, 13*, 31–41.

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: John Wiley.

Cronbach, L. J., Gleser, G. C. and Rajaratnam, N. (1963). Theory of generalizability. A liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology, 16*, 137–173.

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Cronbach, L. J., and Warrington, W. G. (1952). Efficacy of multiple choice tests as a function of spread of item difficulties. *Psychometrika, 17*, 127–147.

Cross, L., and Frary, R. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple choice tests. *Journal of Educational Measurement, 14*, 313–322.

Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement, 10*, 94–96.

Cureton, E. E. (1958). The definition and estimation of test reliability. *Educational Psychological Measurement, 18*, 715–738.

Darlington, R. B. (1970). Some techniques for maximizing a test's validity when the criterion variable is unobserved. *Journal of Educational Measurement, 7,* 1–14.
Darlington, R. B. (1971). Another look at culture fairness. *Journal of Educational Measurement, 8,* 71–82.
Davis, C. E., Hickman, J., and Novick, M. R. (1973). A primer on decision analysis for individually prescribed instruction. (ACT Technical Bulletin No. 17). Iowa City: American College Testing Program.
Davis, F. B. and Fifer, G. (1959). The effect on test reliability and validity on scoring aptitude and achievement tests with weights for every choice. *Educational and Psychological Measurement, 14,* 159–170.
Dawes, R. M. (1972). *Fundamentals of attitude measurement.* New York: John Wiley.
Dennis, W. (Ed). (1948). *Readings in the history of psychology.* New York: Appleton-Century-Crofts.
Diamond, J., and Evans, W. (1973). The correction for guessing. *Review of Educational Research, 43,* 181–191.
Dixon, W. J., Brown, M. B., Engelman, L., Frane, J. W., Hill, M. A., Jennrich, R. I., and Toporek, J. D. (1981). *BMDP statistical software.* Berkeley: University of California Press.
Downey, R. G. (1979). Item-option weighting of achievement tests: Comparative study of methods. *Applied Psychological Measurement, 3,* 453–462.
DuBois, P. (1970). *A history of psychological testing.* Boston: Allyn and Bacon.
Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika, 16,* 407–424.
Ebel, R. L. (1956). Obtaining and reporting evidence on content validity. *Educational and Psychological Measurement, 16,* 269–281.
Ebel, R. L. (1962). Content standard test scores. *Educational and Psychological Measurement, 22,* 15–25.
Ebel, R. L. (1965). *Measuring educational achievement.* Englewood Cliffs, N.J.: Prentice-Hall.
Ebel, R. L. (1968). The value of internal consistency in classroom examinations. *Journal of Educational Measurement, 5,* 71–74.
Ebel, R. L. (1972). *Essentials of educational measurement* (2nd Ed.). Englewood Cliffs, N.J.: Prentice-Hall.
Ebel, R. L. (1982). Proposed solutions to two problems of test construction. *Journal of Educational Measurement, 19,* 267–278.
Echternacht, G. J. (1972). The use of confidence testing in objective tests. *Review of Educational Research, 42,* 217–236.
Echternacht, G. J. (1974). A quick method for determining item bias. *Educational and Psychological Measurement, 34,* 271–280.
Echternacht, G. (1977). Grade-equivalent scores. *Measurement in Education, 8*(2), 1–4.
Englehart, M. D. (1965). A comparison of several item discrimination indices. *Journal of Educational Measurement, 2,* 69–76.
Fahner, S. (1974). Item sampling and decision making in achievement testing. *British Journal of Mathematical and Statistical Psychology, 27,* 172–175.
Fechner, G. T. (1860). Elements of psychophysics. In W. Dennis (Ed.), *Readings in the history of psychology.* New York: Appleton-Century-Crofts, 1948.
Findley, W. G. (1956). A rationale for evaluation of item discrimination statistics. *Educational and Psychological Measurement, 16,* 175–180.

Fishbein, M. (Ed.). (1967). *Readings in attitude theory and measurement.* New York: John Wiley.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327–358.

Forsyth, R., Saisangjan, V., and Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. *Applied Psychological Measurement, 5,* 175–186.

Frederiksen, N. (1981). The real test bias. (Research Report No. 81–40). Princeton, N.J.: Educational Testing Service.

Frick, T., and Semmel, M. I. (1978). Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research, 48,* 157–184.

Frisbie, D. A., and Brandenburg, D. C. (1979). Equivalence of questionnaire items with varying response formats. *Journal of Educational Measurement, 16,* 43–48.

Galton, F. (1883). Inquiries into human faculty and its development. In W. Dennis (Ed). *Readings in the history of psychology.* New York: Appleton-Century-Crofts, 1948.

Ganapole, S. J. (1980). The fundamental reading competencies test. *Journal of Educational Measurement, 7,* 71–73.

Gardner, P. L. (1975). Scales and statistics. *Review of Educational Research, 45,* (1), 43–58.

Gavin, A. T. (1977). Guide to the development of written tests for selection and promotion: The content validity model. (Technical memorandum 77-6). Washington D.C.: U.S. Civil Service Commission.

Gilman, D. A., and Ferry, P. (1972). Increasing test reliability through self-scoring procedures. *Journal of Educational Measurement, 9,* 205–207.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist, 18,* 519–521.

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15,* 237–262.

Gleser, G. C., Cronbach, L. J., and Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika, 30,* 395–418.

Gordon, I. J. (1967). *A test manual for the How I See Myself Scale.* Florida Educational Research and Development Council.

Gross, A. L., and Su, W. (1975). Defining a "fair" or "unbiased" selection model: A question of utilities. *Journal of Applied Psychology, 60,* 345–351.

Guertin, W. H., and Bailey, J. D., Jr. (1970). *Introduction to modern factor analysis.* Ann Arbor, Mich.: Edwards Brothers.

Guilford, J. P. (1954). *Psychometric methods* (2nd Ed.). New York: McGraw-Hill.

Guion, R. M. (1979). Content validity—the source of my discontent. *Applied Psychological Measurement, 1,* 1–10.

Guion, R. M. (1978). "Content validity"—in moderation. *Personnel Psychology, 31,* 205–213.

Guion, R. M. (1980). On trinitarian doctrines of validity. *Professional Psychology, 11,* 385–398.

Gulliksen, H. (1945). The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika, 10,* 79–91.

Gulliksen, H. (1950). *Theory of mental tests.* New York: John Wiley.

Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t. *Annals of Mathematical Statistics, 34,* 792–828.

Gustafsson, J. E. (1980). A solution of the conditional estimation problem for long tests in

the Rasch model for dichotomous items. *Educational and Psychological Measurement, 40*, 377–385.

Guttman, L. (1941a). An outline of the statistical theory of prediction. In P. Horst et al. (Eds.). *The prediction of personal adjustment*. New York: Social Science Research Council.

Guttman, L. (1941b). The quantification of class attributes: A theory and method of scale construction. In P. Horst et al. (Eds.). *The prediction of personal adjustment*. New York: Social Science Research Council.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255–282.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer et al., *Measurement and prediction*. Princeton, N.J.: Princeton University Press.

Guttman, L. (1969). Integration of test design and analysis. *Proceedings of the 1969 invitational conference on testing problems*. Princeton, N.J.: Educational Testing Service.

Hambleton, R. K. (1978). On the use of cutoff scores with criterion-referenced tests in instructional settings. *Journal of Educational Measurement, 15*, 277–290.

Hambleton, R. K. (1980). Test score validity and standard setting methods. In R. A. Berk (Ed.). *Criterion-referenced measurement: The state of the art*. Baltimore: Johns Hopkins University Press.

Hambleton, R. K. (Ed.). (1983). *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

Hambleton, R. K., and Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement, 14*, 75–96.

Hambleton, R. K., and Murray, L. (1983). Some goodness of fit investigations for item response models. In R. K. Hambleton (Ed.). *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

Hambleton, R. K., and Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*, 159–170.

Hambleton, R. K., Swaminathan, J., Algina, J., and Coulson, D. B. (1978a). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research, 48*, 1–47.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., and Gifford, J. A. (1978b). Developments in latent trait theory: Models, technical issues, and applications. *Review of Educational Research, 48*, 467–510.

Hambleton, R. K., and Traub, R. E. (1971). Information curves and efficiency of three logistic test models. *British Journal of Mathematical and Statistical Psychology, 24*, 273–281.

Hanna, G. S. (1975). Incremental reliability and validity of multiple choice tests with an answer-until-correct procedure. *Journal of Educational Measurement, 12*, 175–178.

Harman, H. H. (1967). *Modern factor analysis* (2nd Ed.). Chicago: University of Chicago Press.

Harris, C. W. (1974). Problems of objectives-based measurement. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.). *Problems in criterion referenced measurement*. (CSE monograph series in evaluation No. 3). Los Angeles: Center for the Study of Evaluation, University of California.

Harris, C. W., Alkin, M. C., and Popham, W. J. (1974). *Problems in criterion referenced measurement*. (CSE monograph series in evaluation, No. 3). Los Angeles: Center for the Study of Evaluation, University of California.

Harris, C. W., and Pearlman, A. P. (1977). Conventional significance tests and indices of agreement or association. In C. W. Harris, A. P. Pearlman, and R. R. Wilcox (Eds.). *Achievement test items—Methods of study.* (CSE monograph series in evaluation, No. 6) Los Angeles: Center for the Study of Evaluation, University of California.

Harris, C. W., Pearlman, A. P., and Wilcox, R. R. (1977). *Achievement test items—Methods of study.* (CSE monograph series in evaluation, No. 6). Los Angeles: Center for the Study of Evaluation, University of California.

Hartley, A. A., and Hartley, J. T. (1976). Predicting performance in the basic research methods course in psychology. *Educational and Psychological Measurement, 36,* 449–452.

Hays, W. L. (1981). *Statistics for psychologists.* New York: Holt, Rinehart and Winston.

Hendrickson, G. F. (1971). The effect of differential option weighting on multiple choice objective tests. *Journal of Educational Measurement, 8,* 291–296.

Henryssen, S. (1971). Gathering, analyzing and using data on test items. In R. L. Thorndike (Ed.). *Educational measurement* (2nd Ed.). Washington, D.C.: American Council on Education.

Hilton, T. L., and Rhett, H. (1973). The base year study of the national longitudinal study of the high school class of 1972. (Final Report Contract No. OEC-0-72-0903). Office of Education, National Center for Educational Statistics, U.S. Department of Health, Education, and Welfare. Princeton, N.J.: Educational Testing Service.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics, 32,* 1–49.

Holland, W. F., and Keller, R. T. (1978). A cross validation study of the Kirton Adaptation-Innovation Inventory in three research and development organizations. *Applied Psychological Measurement,* 2, 510–563.

Holmes, S. E. (1982). Unidimensionality and vertical equating with the Rasch model. *Journal of Educational Measurement, 19,* 139–147.

Hoyt, C. J. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6,* 153–160.

Hulin, C. L., Drasgow, F., and Parsons, C. K. (1983). *Item response theory.* Homewood Ill.: Dow Jones-Irwin.

Hulin, C. L., Lissak, R. I., and Drasgow, F. (1982). Recovery of two and three parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement, 6,* 249–260.

Hunter, J. E. (1975). A critical analysis of the use of item means and item test correlations to determine the presence or absence of content bias in achievement test items. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, Md.

Hunter, J. E., and Cohen, S. H. (1974). Correcting for unreliability in nonlinear models of attitude change. *Psychometrika, 34,* 445–468.

Hunter, J. E., and Schmidt, F. L. (1976). A critical analysis of the statistical and ethical implications of various definitions of "test bias." *Psychological Bulletin, 83,* 1,053–1,071.

Hunyh, H. (1976a). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement, 13,* 253–264.

Hunyh, H. (1976b). Statistical considerations of mastery scores. *Psychometrika, 41,* 65–78.

Hunyh, H. (1980a). Computation and inference for two reliability indices in mastery testing based on the beta-binomial model. In H. Hunyh and J. C. Saunders. Solutions for some technical problems in domain-referenced mastery testing. (Final Report, Project No. NIE-

G-78-0087). National Institute of Education, Department of Health, Education, and Welfare.

Hunyh, H. (1980b). A nonrandomized minimax solution for passing scores in the binomial error model. *Psychometrika, 45*, 167–182.

Hunyh, H. (1980c). A note on decision theoretic coefficients for tests. In H. Hunyh and J. C. Saunders. Solutions for some technical problems in domain-referenced mastery testing. (Final Report, Project No. NIE-G-78-0087). National Institute of Education, Department of Health, Education, and Welfare.

Hunyh, H. (1980d). Statistical inference for false positive and false negative error rates in mastery testing (computer program and tables added). In H. Hunyh and J. C. Saunders. Solutions for technical problems in domain-referenced mastery testing. (Final Report, Project No. NIE-G-78-0087). National Institute of Education, Department of Health, Education, and Welfare.

Hunyh, H., and Saunders, J. C. (1980). Accuracy of two procedures for estimating reliability of mastery tests. *Journal of Educational Measurement, 17*, 351–358.

Jennrich, R. I., and Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika, 31*, 313–323.

Ironson, G. H. (1982). Use of chi-square and latent trait approaches for detecting item bias. In R. A. Berk (Ed.). *Handbook of methods for detecting item bias.* Baltimore: Johns Hopkins University Press.

Ironson, G., Homan, S., Willis, R., and Signer, B. (In press). The validity of item bias techniques with math word problems. *Applied Psychological Measurement.*

Ironson, G. H., and Subkoviak, M. (1979). A comparison of several methods for assessing item bias. *Journal of Educational Measurement, 16*, 209–225.

Irvin, L. K., Halpern, A. S., and Landman, J. T. (1980). Assessment of retarded student achievement with standardized true/false and multiple-choice tests. *Journal of Educational Measurement, 17*, 51–58.

Jaeger, R. M. (1979). Measurement consequences of selected standard-setting models. In M. A. Bunda, and J. R. Sanders (Eds.). *Practices & problems in competency based measurement.* National Council on Measurement in Education.

Jaeger, R. M. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis, 4*, 461–475.

Jaeger, R. M. (1984). *Sampling in education and the social sciences.* New York: Longman.

Jensen, A. R. (1980). *Bias in mental testing.* New York: The Free Press.

Jersild, A. J. (1952). *In search of self.* New York: Teachers College Bureau of Publications.

Joncich, G. M. (1968). *The sane positivist: A biography of Edward L. Thorndike.* Middletown, Conn.: Wesleyan University Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.

Jöreskog, K. G., and Sorbom, D. (1984). LISREL VI, Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods. Mooresville, Ind.: Scientific Software, Inc.

Kaiser, H. G. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.

Kane, M. T. (1982). A sampling model for validity. *Applied Psychological Measurement, 6*, 125–160.

Kane, M. T., and Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement, 4*, 105–126.
Kane, M. T., Gillmore, G. M., and Crooks, T. J. (1976). Student evaluations of teaching: The generalizability of class means. *Journal of Educational Measurement, 13*, 171–184.
Kaplan, R. M., and Saccuzzo, D. P. (1982). *Psychological testing principles, applications, and issues*. Monterey, Calif.: Brooks/Cole.
Katz, M. (1958). Selecting an achievement test: Principles and procedures. In V. H. Noll, D. P. Scannell, and R. P. Noll (Eds.). *Introductory readings in educational measurement*. Boston: Houghton Mifflin, 1972.
Keats, J. A., and Lord, F. M. (1962). A theoretical distribution for mental test scores. *Psychometrika, 27*, 215–231.
Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: World Book.
Kelley, T. L. (1939). Selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17–24.
Kelley, T. L. (1942). The reliability coefficient. *Psychometrika, 7*, 75–83.
Kiefer, J., and Wolfowitz, J. (1956). Consistency of maximum likelihood estimates in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics, 27*, 887–890.
Kish, L. (1965). *Survey sampling*. New York: John Wiley.
Klein, S. P., and Kosecoff, J. P. (1975). Determining how well a test measures your objectives. (CSE Report No. 94). Los Angeles: Center for the Study of Evaluation, University of California.
Kleinbaum, D. G., and Kupper, L. L. (1978). *Applied regression analysis and other multivariable methods*. North Scituate, Mass.: Duxbury Press.
Koffler, S. L. (1980). A comparison of approaches for setting standards. *Journal of Educational Measurement, 17*, 167–178.
Kolen, M. J. (1984). Effectiveness of analytical smoothing in equipercentile equating. *Journal of Educational Statistics, 9*, 25–44.
Krathwohl, D. R., Bloom, B. S., and Masia, B. (1964). *Taxonomy of educational objectives, handbook II: The affective domain*. New York: McKay.
Kuder, G. F., and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151–160.
Kurtz, A. K., and Mayo, S. T. (1979). *Statistical methods in education and psychology*. New York: Springer-Verlag.
Lam, T. C. M., and Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*, 317–322.
Lange, A., Mehrens, W. A., and Lehmann, I. J. (1967). Using item analysis to improve tests. *Journal of Educational Measurement, 4*, 65–68.
Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinborough, 6*, 73–287.
Lawley, D. N., and Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd Ed.). New York: American Elsevier Publishing Company.
Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563–575.
Levine, M. V. (1981). Weighted item bias statistics. (Report 81-5). Urbana-Champaign: Department of Educational Psychology, University of Illinois.
Levine, M. V., Wardrop, J. L., and Linn, R. L. (1982). Weighted mean square item bias statistics. Paper presented at the annual meeting of the American Educational Research Association, New York.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, (140), 44–53.
Lindquist, E. F. (1936). The theory of test construction. In H. E. Hawkes, E. F. Lindquist, and C. Mann (Eds.). *The construction and use of achievement examinations*. Boston: Houghton Mifflin.
Lindquist, E. F. (1953). *Design and analysis of experiments in education and psychology*. Boston: Houghton Mifflin.
Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43*, 343–357.
Linn, R. L., Levine, M. V., Hastings, C. N., and Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement, 5*, 159–173.
Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. *Journal of Educational Measurement, 9*, 13–26.
Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monograph, 61*, No. 4.
Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin, 51*, 493–504.
Lomax, R. G. and Algina, J. (1979). Comparison of two procedures for analyzing multitrait multimethod matrices. *Journal of Educational Measurement, 16*, 177–186.
Lord, F. M. (1952a). The relationship of the reliability of multiple choice items to the distribution of item difficulties. *Psychometrika, 18*, 181–194.
Lord, F. M. (1952b). A theory of test scores. *Psychometric Monograph*, No. 7.
Lord, F. M. (1953). The relationship of test score to trait underlying the test. *Educational and Psychological Measurement, 13*, 517–548.
Lord, F. M. (1955). Sampling fluctuations resulting from the sampling of test items. *Psychometrika, 20*, 1–22.
Lord, F. M. (1957). Do tests of the same length have the same standard error of measurement? *Educational and Psychological Measurement, 17*, 510–521.
Lord, F. M. (1959a). Test norms and sampling theory. *Journal of Experimental Education, 27*, 247–263.
Lord, F. M. (1959b). Tests of the same length do have the same standard error of measurement. *Educational and Psychological Measurement, 19*, 233–239.
Lord, F. M. (1963). Formula scoring and validity. *Educational and Psychological Measurement, 23*, 663–672.
Lord, F. M. (1965). A strong true-score theory with application. *Psychometrika, 30*, 239–270.
Lord, F. M. (1968). An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement, 28*, 989–1020.
Lord, F. M. (1971). The self scoring flexilevel test. *Journal of Educational Measurement, 8*, 147–151.
Lord, F. M. (1975). Formula scoring and number-right scoring. *Journal of Educational Measurement, 12*, 7–12.
Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: Lawrence Erlbaum.
Lord, F. M. (1982). The standard error of equipercentile equating. *Journal of Educational Statistics, 1*, 165–192.
Lord, F. M., and Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.

Loyd, B. H., and Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement, 17,* 179–193.

McCall, W. A. (1939). *Measurement.* New York: Macmillan.

McClung, M. S. (1978). Developing proficiency programs in California public schools: Some legal implications and a suggested implementations schedule. Sacramento: California State Department of Education.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34,* 100–117.

McNemar, Q. (1962). *Psychological statistics.* New York: John Wiley.

McNemar, Q. (1975). On so-called test bias. *American Psychologist, 30,* 848–851.

Magnusson, D. (1967). *Test theory.* Boston: Addison-Wesley.

Marascuilo, L. A., and McSweeney, M. (1977). *Nonparametric and distribution free methods for the social sciences.* Monterey, Calif.: Brooks/Cole.

Marco, G. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14,* 139–160.

Marsh, H. W., and Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20,* 231–248.

Marx, R. W., and Winne, P. H. (1978). Construct interpretations of three self-concept inventories. *American Educational Research Journal, 15,* 99–109.

Masters, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement, 11,* 49–53.

Matarazzo, J. D., Allen, B. V., Soslow, G., and Weins, A. N. (1964). Characteristics of successful policemen and firemen applicants. *Journal of Applied Psychology, 48,* 123–133.

Medley, D. M., and Meitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage, *Handbook of research on teaching.* Chicago: Rand McNally.

Mehrens, W. A., and Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology* (3rd ed.). New York: Holt, Rinehart, and Winston.

Meskauskas, J. A. (1976). Evaluation models for criterion-referenced testing: Views regarding mastery and standard setting. *Review of Educational Research, 46,* 133–158.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30,* 955–966.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*(9), 9–20.

Millman, J. (1973). Passing scores and test lengths for criterion-referenced tests. *Review of Educational Research, 43,* 205–216.

Millman, J. (1974). Criterion referenced measurement. In W. J. Popham (Ed.). *Evaluation in education: Current applications.* Berkeley, Calif.: McCutcheon.

Millman, J., and Glass, G. V. (1967). Rules of thumb for writing the ANOVA table. *Journal of Educational Measurement, 4,* 41–51.

Millman, J., and Popham, W. J. (1974). The issue of item and test variance for criterion-referenced tests: A reply. *Journal of Educational Measurement, 11,* 137–138.

Mills, C. N. (1983). A comparison of three methods of establishing cutoff scores on criterion referenced tests. *Journal of Educational Measurement, 20,* 283–292.

Morrison, D. F. (1976). *Multivariate statistical methods* (2nd Ed). New York: McGraw-Hill.

Mosher, D. L. (1968). Measurement of guilt in females by self-report inventories. *Journal of Consulting and Clinical Psychology, 32,* 690–695.

Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educational and Psychological Measurement, 7,* 191–206.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw-Hill.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd Ed.). Boston: Allyn and Bacon.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement, 14,* 3–19.

Novick, M. R., and Lewis, C. (1974). Prescribing test length for criterion referenced measurement. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.). *Problems in criterion-referenced measurement.* Los Angeles: Center for the Study of Evaluation, University of California.

Novick, M. R., and Lindley, D. V. (1978). The use of more realistic utility functions in educational applications. *Journal of Educational Measurement, 15,* 181–191.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Osgood, C., Suci, G., and Tannenbaum, P. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Oosterhof, A. C. (1976). Similarity of various item discrimination indices. *Journal of Educational Measurement, 13,* 145–150.

Overall, J. E., and Klett, C. J. (1972). *Applied multivariate analysis.* New York: McGraw-Hill.

Patnaik, D., and Traub, R. E. (1973). Differential weighting by judged degree of correctness. *Journal of Educational Measurement, 10,* 281–286.

Pearson, K. (1909). On a new method of determining a correlation between a measured character of A and a character of B, of which only the percentage of cases wherein B exceeds (or falls short of) intensity is recorded for each grade of A. *Biometrika, 7,* 96–105.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd Ed.). York: Holt, Rinehart, and Winston.

Peng, C. J., and Subkoviak, M. J. (1980). A note on Hunyh's normal approximation procedure for estimating criterion-referenced reliability. *Journal of Educational Measurement, 17,* 359–368.

Petersen, N. S. (1976). An expected utility model for "optional" selection. *Journal of Educational Statistics, 1,* 333–358.

Petersen, N. S. (1979). Review of *Achievement test items: Methods of study* by Chester W. Harris, Andrea Pastorak Pearlman, and Rand R. Wilcox. *Journal of Educational Measurement, 16,* 137–138.

Petersen, N. S., and Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement, 13,* 3–29.

Piers, E. V., and Harris, D. B. (1964). Age and other correlates of self-concept. *Journal of Educational Psychology, 55,* 91–96.

Plake, B. S., and Hoover, H. D. (1977). An analytic method of identifying biased test items. Paper presented at the annual meeting of the American Educational Research Association, New York.

Popham, W. J. (1974). Selecting objectives and generalizing test items. In C. W. Harris, M. C. Alkin, and W. J. Popham (Eds.). *Problems in criterion-referenced measurement.* (pp. 13–25). Los Angeles: Center for the Study of Evaluation, University of California.

Popham, W. J. (1978). As always, provocative. *Journal of Educational Measurement, 15,* 297–300.

Popham, W. J. (1981). *Modern educational measurement*. Englewood Cliffs, N.J.: Prentice-Hall.

Popham, W. J., and Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6*, 1–9.

Pough, F. H. (1960). *A field guide to rocks and minerals*. Boston: Houghton Mifflin.

Powers, D. E. (1982). Long term predictive and construct validity of two traditional predictors of law school performance. *Journal of Educational Psychology, 74*, 568–576.

Powers, S., Slaughter, H., and Helmick, C. (1983). A test of the equipercentile hypothesis of the TIERS norm-referenced model. *Journal of Educational Measurement, 20*, 299–302.

Prescott, G. A., Balow, I. H., Hogan, T. P., and Farr, R. C. (1978). *Metropolitan Achievement Tests*. New York: The Psychological Corporation.

Prien, E. P. (1977). The function of job analysis in content validation. *Personnel Psychology, 30*, 1977, 167–173.

Rajaratnam, N. (1960). Reliability formulas for independent decision data when reliability data are matched. *Psychometrika, 25*, 261–271.

Rajaratnam, N., Cronbach, L. J., and Gleser, G. C. (1963). Generalizability of stratified parallel tests. *Psychometrika, 30*, 39–56.

Ree, M. J., and Jensen, H. E. (1980). Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss (Ed.). *Proceedings of the 1979 computerized adaptive testing conference*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.

Rentz, R. R., and Bashaw, W. L. (1977). The national reference scale for reading: An application of the Rasch model. *Journal of Educational Measurement, 14*, 161–180.

Roid, G., and Haladyna, T. (1980). The emergence of an item writing technology. *Review of Educational Research, 50*, 293–314.

Roth, A. V., and Lubin, B. (1981). Factors underlying the depression adjective check lists. *Educational and Psychological Measurement, 41*, 382–385.

Rovinelli, R. J., and Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 2*, 49–60.

Rowley, G. L., and Traub, R. E. (1977). Formula scoring, number-right scoring, and test-taking strategy. *Journal of Educational Measurement, 14*, 15–22.

Rozeboom, W. W. (1981). The cross validational accuracy of sample regressions. *Journal of Educational Statistics, 6*, 179–198.

Rudner, L. M. (1977). An approach to biased item identification using latent trait measurement theory. Paper presented at the annual meeting of the American Educational Research Association, New York.

Rudner, L. M., Getson, P. R., and Knight, D. L. (1980). A Monte Carlo comparison of seven biased item detection techniques. *Journal of Educational Measurement, 17*, 1–10.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review, 9*, 99–103.

Saunders, J. C., Ryan, J. P., and Hunyh, H. (1981). A comparison of two approaches to standard setting based on the Nedelsky procedure. *Applied Psychological Measurement, 5*, 209–218.

Saupe, J. L. (1966). Selecting items to measure change. *Journal of Educational Measurement, 3*, 223–226.

Sawyer, R. L., Cole, N. S., and Cole, J. W. L. (1976). Utilities and the issue of fairness in a decision theoretic model for selection. *Journal of Educational Measurement, 13*, 59–76.

Sax, G. (1980). *Principles of educational and psychological measurement and evaluation* (2nd Ed.). Belmont, Calif.: Wadsworth.

Scheuneman, J. (1979). A new method for assessing bias in test items. *Journal of Educational Measurement, 16*, 143–152.

Schmidt, F. L., and Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36*, 1,128–1,137.

Schmidt, F. L., Hunter, J. E., and Urry, V. W. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology, 61*, 473–485.

Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 2*, 157–174.

Scriven, M. (1978). How to anchor standards. *Journal of Educational Measurement, 15*, 273–275.

Shavelson, R. J. (1981). *Statistical reasoning for the behavioral sciences*. Boston: Allyn and Bacon.

Shavelson, R. J., and Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology, 74*, 3–17.

Shepard, L. A. (1976). Setting standards and living with them. *Florida Journal of Educational Research, 18*, 23–32.

Shepard, L. A. (1979). Setting standards. In M. A. Bunda and J. R. Sanders (Eds.). *Practices and problems in competency-based measurement*. National Council of Measurement in Education.

Shepard L. (1980). Standard setting issues and methods. *Applied Psychological Measurement, 4*, 447–465.

Shepard, L., Camilli, G., and Averill, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criterion. *Journal of Educational Statistics, 6*, 317–376.

Shoemaker, D. M. (1975). Toward a framework for achievement testing. *Review of Educational Research, 45*, 127–148.

Sirotnik, K. (1972). Estimates of coefficient alpha for finite populations of items. *Educational and Psychological Measurement, 32*, 129–136.

Slakter, M. J. (1968). The penalty for not guessing. *Journal of Educational Measurement, 5*, 141–144.

Slakter, M. J. (1969). Generality of risk taking on objective examinations. *Educational and Psychological Measurement, 29*, 125–128.

Slinde, J. A., and Linn, R. L. (1978). An exploration of the adequacy of the Rasch model for vertical equating. *Journal of Educational Measurement, 15*, 23–35.

Slinde, J. A., and Linn, R. L. (1979). The Rasch model, objective measurement, equating, and robustness. *Applied Psychological Measurement, 3*, 437–456.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161–169.

Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology, 5*, 417–426.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.). *Educational measurement* (2nd Ed.). (pp. 359–442). Washington, D.C.: American Council on Education.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement, 13*, 265–275.

Subkoviak, M. J. (1978). Empirical investigation of procedures for estimating reliability for mastery tests. *Journal of Educational Measurement, 15*, 111–116.

Subkoviak, M. J. (1980). Decision consistency approaches. In R. A. Berk (Ed.). *Criterion-referenced measurement: The state of the art*. Baltimore: Johns Hopkins University Press.

Swaminathan, H. (1983). Parameter estimation in item response models. In R. K. Hambleton (Ed.). *Applications of item response theory*. Vancouver: Educational Research Institute of British Columbia.

Swaminathan, H., and Gifford, J. (1979). Estimation of parameters in latent trait models. In D. Weiss (Ed.). *Proceedings of the 1979 computerized adaptive testing conference*. Minneapolis: University of Minnesota.

Swaminathan, H., Hambleton, R. K., and Algina, J. (1974). Reliability of criterion referenced tests: A decision theoretic formulation. *Journal of Educational Measurement, 11*, 263–268.

Tallmadge, G. K., and Wood, C. T. (1976). User's guide: ESEA Title I evaluation and reporting system. Mountain View, Calif.: RMC Research Corp.

Taylor, H. C., and Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565–578.

Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology, 30*, 47–54.

Thissen, D. M. (1976). Information in wrong responses to the Raven Progressive Matrices. *Journal of Educational Measurement, 13* (3), 201–214.

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York: Science Press.

Thorndike, R. L. (1949). *Personnel selection*. New York: John Wiley.

Thorndike, R. L. (1971). Concepts of culture fairness. *Journal of Educational Measurement, 8*, 63–70.

Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4*(5), 3–7.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

Thorndike, R. L., and Hagen, E. (1977). *Measurement and evaluation in psychology and education*. New York: John Wiley.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529–554.

Thurstone, L. L. (1942). *Multiple factor analysis*. Chicago: University of Chicago Press.

Thurstone, L. L., and Chave, E. J. (1929). *The measurement of attitude*. Chicago: University of Chicago Press.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley.

Traub, R. E., and Hambleton, R. K. (1972). The effect of scoring instructions and degree of speededness on the validity and reliability of multiple-choice tests. *Educational and Psychological Measurement, 32*, 737–758.

Traub, R. E., and Rowley, G. L. (1980). Reliability of test scores and decisions. *Applied Psychological Measurement, 4*, 517–546.

Traub, R. E., and Wolfe, R. G. (1981). Latent trait theories and assessment of educational achievement. In D. C. Berliner (Ed.). *Review of research in education 9*. Washington, D.C.: American Educational Research Association.

Travers, R. M. (1951). Rational hypotheses in the construction of tests. *Educational and Psychological Measurement, 11*, 128–137.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 11*, 1–13.
Udinsky, B. F., Osterlind, S. J., and Lynch, S. W. (1981). *Evaluation resource handbook: Gathering, analyzing, reporting data*. San Diego, Calif.: EDITS.
van der Linden, Wim J. (1982). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard-setting. *Journal of Educational Measurement, 19*, 295–380.
van der Linden, W. J., and Mellenbergh, G. J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement, 1*, 593–599.
van der Linden, W. J., and Mellenbergh, G. J. (1978). Coefficients for tests from a decision theoretic point of view. *Applied Psychological Measurement, 2*, 119–134.
van der Ven, A. H. G. S. (1980). *Introduction to scaling*. New York: John Wiley.
Velicer, W. F., and Stevenson, J. F. (1978). The relation between item format and the structure of the Eysenck Personality Inventory. *Applied Psychological Measurement, 2*, 293–304.
Wainer, H., Morgan, A., and Gustafsson, J. E. (1980). A review of estimation procedures for the Rasch model with an eye toward longish tests. *Journal of Educational Statistics, 5*, 35–64.
Ward, W. C., Carlson, S. B., and Woisetschlager, E. (1983). Ill-structured problems as multiple-choice items. (GRE Board Professional Report No. 81–18P). Princeton, N.J.: Educational Testing Service.
Ward, W. C., Frederiksen, N., and Carlson, S. B. (1980). Construct validity of free-response and machine-scorable forms of a test. *Journal of Educational Measurement, 17*, 11–29.
Ware, W. B., and Benson, J. (1975). Appropriate statistics and measurement scales. *Science Education, 59* (4), 575–582.
Weber, E. H. (1846). The sense of touch and common feeling. In W. Dennis (Ed). *Readings in the history of psychology*. New York: Appleton-Century-Crofts, 1948.
Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams and Wilkins.
Wechsler, S. W. (1958). *The measurement and appraisal of adult intelligence* (4th Ed.). Baltimore: Williams and Wilkins.
Weitzenhoffer, A. M. (1951). Mathematical structures and psychological measurement. *Psychometrika, 16*, 387–406.
Whitely, S. E., and Dawis, R. V. (1974). The nature of objectivity with the Rasch model. *Journal of Educational Measurement, 11*, 163–178.
Wilcox, R. R. (1976). A note on the length and passing score of a mastery test. *Journal of Educational Statistics, 1*, 359–364.
Wilcox, R. R. (1977a). Estimating the likelihood of false positive and false negative decisions in mastery decisions: An empirical Bayes approach. *Journal of Educational Statistics, 2*, 289–307.
Wilcox, R. R. (1977b). New methods for studying equivalence. In C. W. Harris, A. P. Pearlman, and R. R. Wilcox (Eds.). *Achievement test items—New methods of study*. Los Angeles: Center for the Study of Evaluation, University of California.
Wilcox, R. R. (1978). A note on decision theoretic coefficients for tests. *Applied Psychological Measurement, 2*, 609–613.
Wilcox, R. R. (1981a). A closed sequential procedure for answer-until-correct tests. *Journal of Experimental Education, 5*, 219–222.
Wilcox, R. R. (1981b). A cautionary note on estimating the reliability of a mastery test with the beta-binomial model. *Applied Psychological Measurement, 5*, 531–537.

Wilcox, R. R. (1981c). Solving measurement problems with an answer-until-correct procedure. *Applied Psychological Measurement, 5*, 399–414.

Wilcox, R. R. (1982). Some new results on an answer-until-correct procedure. *Journal of Educational Measurement, 19*, 67–74.

Wood, R. L., Wingersky, M. S., and Lord, F. M. (1976). LOGIST-A computer program for estimating examinee ability and item characteristic curve parameters. (Research Memorandum 76-6). Princeton, N.J.: Educational Testing Service.

Wright, B. D. (1968). Sample free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service.

Wright, B. D., and Douglas, G. A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement, 37*, 47–60.

Wright, B. D., Mead, R., and Bell, S. (1979). *BICAL: Calibrating items with the Rasch model*. (Statistical Laboratory Department of Education RM 23b). Chicago: University of Chicago.

Wright, B. D., and Panchapakesan, N. (1969). A procedure for sample free item analysis. *Educational and Psychological Measurement, 29*, 23–48.

Wright, B. D., and Stone, M. (1979). *Best test design*. Chicago: MESA Press.

Wundt, W. (1873). Principles of physiological psychology. In W. Dennis (Ed.). *Readings in the history of psychology*. New York: Appleton-Century-Crofts, 1948.

Yalow, E. S., and Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher, 12*(8), 10–14.

Yerkes, R. M. (1921). Psychological examining in the United States Army. In W. Dennis (Ed.). *Readings in the history of psychology*. New York: Appleton-Century-Crofts, 1948.

Zieky, M. J., and Livingston, S. A. (1977). Manual for setting standards on the Basic Skills Assessment Tests. Princeton, N.J.: Educational Testing Service.

Zimmerman, D. W., and Williams, R. H. (1982). Gain scores in research can be highly reliable. *Journal of Educational Measurement, 19*, 149–154.

الملحق (أ)

Probabilities and *Y* Ordinates Associated with given *z*-Score under the Standard Normal Curve

Absolute Value of *z*-Score	$Pr \leq -z$	$Pr \leq +z$	*Y* ordinate
.00	.500	.500	.3989
.05	.480	.520	.3984
.10	.460	.540	.3970
.15	.440	.560	.3945
.20	.421	.579	.3910
.25	.401	.599	.3867
.30	.382	.618	.3814
.35	.363	.637	.3752
.40	.344	.656	.3683
.45	.326	.674	.3605
.50	.309	.691	.3521
.55	.291	.709	.3429
.60	.274	.726	.3332
.65	.258	.742	.3230
.70	.242	.758	.3123
.75	.227	.773	.3011
.80	.212	.788	.2897
.85	.198	.802	.2780
.90	.184	.816	.2661
.95	.171	.829	.2541
1.00	.159	.841	.2420
1.05	.147	.853	.2299
1.10	.136	.864	.2179
1.15	.125	.875	.2059
1.20	.115	.885	.1942
1.25	.106	.894	.1827
1.30	.097	.903	.1714
1.35	.089	.911	.1604
1.40	.081	.919	.1497
1.45	.074	.926	.1394
1.50	.067	.933	.1295
1.55	.061	.939	.1200
1.60	.055	.945	.1109
1.65	.049	.951	.1023
1.70	.045	.955	.0941
1.75	.040	.960	.0863
1.80	.036	.964	.0790
1.85	.032	.968	.0721
1.90	.029	.971	.0656
1.95	.026	.974	.0596
2.00	.023	.977	.0540
2.05	.020	.980	.0488

تابع ملحق (أ)

Absolute Value of z-Score	$Pr \leq -z$	$Pr \leq +z$	Y ordinate
2.10	.018	.982	.0440
2.15	.016	.984	.0396
2.20	.014	.986	.0355
2.25	.012	.988	.0317
2.30	.011	.989	.0283
2.35	.009	.991	.0252
2.40	.008	.992	.0224
2.45	.007	.993	.0198
2.50	.006	.994	.0175
2.55	.005	.995	.0155
2.60	.005	.995	.0136
2.65	.004	.996	.0120
2.70	.003	.997	.0104
2.75	.003	.997	.0091
2.80	.003	.997	.0079
2.85	.002	.998	.0069
2.90	.002	.998	.0060
2.95	.002	.998	.0051
3.00	.001	.999	.0044

الملحق (ب)

مفاتيح إجابة الأسئلة

الفصل الأول

/1 أ- اختبار

ب- قياس

ج- لا توجد طريقة معيارية لجمع البيانات لذا فإنه لا يعد قياس أو اختبار.

د- اختبار ومقياس

هـ- اختبار

و- قياس

/2 أ- قياس الأداء النموذجي

ب- قياس الأداء الملاحظ

ج- قياس الأداء النموذجي

د- قياس أقصى أداء

/3 1- المشكلة الرئيسية هي تحديد بناء الطلاقة وتحديد طريقة القياس. فقد عرف الباحثون البناء إجرائياً بطريقة تختلف عن الطريقة التي اختارها هذا الباحث.

2- أخطاء المعاينة تعد المشكلة الرئيسية للتعميم من أداء العينة من الفقرات المستخدمة إلى مجال أكبر من النطاق. فعند استخدام عينة أخرى من الفقرات فقد تختلف درجات التلاميذ.

3- بالإضافة إلى أخطاء التلاميذ في بعض المواقف التي قد لا تظهر ثانية عند إعادة الاختبار.

4- مشكلة أخرى هي تقرير عدد الدرجات المقدرة لكل إجابة صحيحة.

5- إعطاء الدرجات النهائية على الاختبار لها علاقة نوعاً ما بسلوكات أخرى تمثل الطلاقة التي قد تكون مهمة.

4/ أ – بينية

ب– فوندت وويبر

ج– جالتون

5/ أ– تصميم بحث

ب– تطوير أداه.

ج– تحليل إحصائي استدلالي

د– معاينة.

6/ تطوير اختبار مؤلف من 50 فقرة.

الفصل الثاني

1/ أ– المجموعة الثانية

ب– المجموعة الثانية / المجموعة الأولى

ج– المجموعة الأولى

د– المجموعة الثانية والتي لها منحنى توزيع ملتو قليلاً.

2/ $\mu_1 = 25$ $\sigma_1 = 1.378$ أو 1.70

$\sigma_1^2 = 1.90$ (استخدمت N في المقام في حساب الانحراف المعياري والتباين).

$\mu_2 = 26$ $= 1.703$ أو 1.70

$\sigma_1^2 = 2.89$

$\mu_3 = 25$ $\sigma_{13} = 1.679$ أو 1.67

$\sigma_3 = 2.79$

3/

	الدرجة الخام	الدرجة المنحرفة	الدرجة الزائية
جون	22	2-	2.17-
بيتر	28	3	2.17
ادوارد	27	2	1.45
كاثي	24	1-	0.72-

/4 أ- لكل Z = صفر، الدرجة الخام هي متوسط كل مجموعة

ب- لكل Z = -1، تحسب من المعادلة $Z = \frac{x - \bar{x}}{r}$

للمجموعة الأولى $-1.00 = \frac{25 - x}{1.38}$

$x = 23.62 \cong 24$

للمجموعة الثانية $x = 24.3 \cong 24$

للمجموعة الثالثة $x = 23.33 \cong 23$

/5 لا. لأن تباين المجموعة ٣ أكبر لكل درجة زائية أكبر من صفر تكون الدرجات الخام للمجموعة الثالثة أكبر من المجموعة الأولى.

/6 أ- 1.6 % درجاتهم أقل من ريبيكا.

ب- 1.6 % درجاتهم أعلى من شارون.

ج- 5.1 % درجاتهم بين رونالد وشارون.

د- درجة شيلدون لأنهما الأقرب من وسط التوزيع.

/7 لا. لأن الدرجات توزيعها ليس طبيعياً.

/8 أ- $z_1 = \frac{55 - 50}{10} = 5$

$z = \frac{45 - 50}{10} = -5$ المساحة تحت المنحنى بين z_1 و z_2

$= 0.691 - 0.309$ أو 0.382

ب- سؤال مكافئ: ما نسبة المفحوصين التي تقع درجاتهم بين 45 و 50.

/9 الاستعداد والاستيعاب القرائي، العلاقة موجبة

/10 $\rho_{12} = 0.097$

$\rho_{13} = 0.07$

$\rho_{14}^2 = 0.42$

$\rho_{12}^2 = 0.94$

$\rho_{14} = 0.18$

/11 لا علاقة بين متغيرين اثنين تشير إلى أن أي منهما يكون سبباً لحدوث الآخر.

/12 $(Zy) \quad (\rho_{xy}) = zy_1$

$0.32- = (0.75-)\ (0.42) = zy_1$

$\frac{20 - y\backslash}{7.62} = \frac{\mu\ y - y}{\sigma\ y} = 0.32-$

$17.56 = Y\backslash$ ◄═══

/13 الاختلاف في حجوم العينات قد يكون المؤدي إلى الاستنتاجات المختلفة، والخطأ المعياري للارتباط يمكن حسابه من الصيغة: $\frac{1}{\sqrt{1 - N}}$

/14 نعم ، لا

$\sigma_{y.x} = \sqrt{Pxy - 1\sigma y} = \sqrt{(0.63) - 1} \zeta$ 2.33 بنسبة ثقة 95 % بأن درجات الاختبار البعدي لـ $X = 15$ تقع في المنطقة:

(15 - 1.96 (2.33)) و (15 + 1.96 (2.33)) أو بين (10.43) و (19.75).

/15 $0.63 = byx\ (\frac{3}{5}) = 0.38$ للمفحوصي الأول الذي يبعد نقطة عن X.

$5\ x\ 0.38 = 1.9$ للمفحوص الذي يبعد خمس نقاط عن X.

/16 الميل $byx\ \frac{pxy - \sigma y}{\sigma} = 0.63\ x \frac{3}{5} = 0.038$

القاطع $C = M_y - by.x\ (\mu_x)$

$= 36 - 0.38\ (20) = 28.4$

/17 أ- الوسط= $5\ (10) = 50$

ب- الدرجة المنحرفة لكل مفحوص ستكون خمسة أضعاف

ج- التباين سيكون $5\ x\ 9^2 = 255$ نقطة.

د- الارتباط لا يتغير.

/18 $\sigma_x^{12} = \frac{\zeta (x\backslash - \mu_x\backslash)^2}{N}$

$= \frac{\zeta (xk - \mu_{xk})^2}{N}$

$= \frac{k^2 \zeta (x - \mu_x)^2}{N}$

$\sigma^2\ xk^2$

الفصل الثالث

١/ أ- تحويل خطي $y = 2x + 1$ تدريج رتبي

ب- تدريج اسمي لا يخضع لنظام معين.

ج- مئوي.

٢/ أ- 7 ب- 5 ج-4 د-2

ب- الأفراد 2 7 8 10

ج- 3 : 1 , 4 : 5 , 6 : 9

د- لـS_2 يستجيب لـ ب

S_7 يستجيب لـ ب أو جـ

S_8 يستجيب لـ أ

S_{10} يستجيب لـ ب أو أ

يبدو أن ب هي المشكلة، فعلى مطور الاختبار أن يتساءل عما إذا كانت ب تقيس البعد الذي تقيسه الفقرات الأخرى.

هـ- $C = 1 - \frac{\text{مجموع الأخطاء}}{\text{مجموع الاستجابات}}$

$C - 1 \frac{4}{40} = 1 - 0.1 = 0.90$ هذا التدريج يحقق تدريج جوتمان.

و- من الصعوبة بمكان وعملياً تطوير تدريج لعدد كبير من الفقرات بمثل هذه الطريقة.

/4 أ - أ ب ج د منطقة *1*

ب ج أ د منطقة *3*

ج ب د أ منطقة *5*

ج د ب أ منطقة *6*

ب أ ج د منطقة *2*

د ج ب أ منطقة *7*

ب ج د أ منطقة *4*

/5 أ- متمركز حول المثير

ب- متمركز حول الفرد

ج- متمركز حول الاستجابة.

الفصل الرابع

/1

	توزيع المعالم	توزيع طبيعي	ارتباط / انحدار
المعرفة	أ	ج	ج % 10
الفهم	أ	هـ	ز ط % 30
التطبيق	ب	و	ح ي ك ل % 60
	%35	%15	%50

أ- %10 %30 % 60

ب- %35 % 50

ج- % 30

/4 أ- يتغير كل من درجة حرارة وضغط الغاز.

ب- أن يخبر التلاميذ بعضهم البعض عن كيفية حل المسألة، وهنا فإنها لا تقيس مستوى التطبيق .

/5 أ- استخدام " سبب".

ب- استخدام " إذا ...... فإن".

ج- هذه أساساً حقيقة، لذا من الأفضل استخدام نمط صح ـــ خطأ بدلاً من نمط الاتجاه.

د- استخدام " إلا".

هـ- طويلة جداً، اللفظ معقد جداً. استخدم لفظ " أحياناً".

/6 من الطرائق الممكنة استخدام لفظ التلميذ نفسه في تحليل المحتوى والمحاضرة

والمراجعة والأحداث المهمة والمكتوبة من قبل المعلم أو من أحد المتخصصين المهتمين بالطفولة. (ويمكن استخدام أراء الخبراء والملاحظين ولكنها أقل جدوى مما سبق).

الفصل الخامس

/1 أ- $\mu_1 = 7$ ، $\sigma_1 = 1.14$ (وحسبت باستخدام قيمة N في المقام).

$\mu_2 = 6$ ، $\sigma_2 = 1.90$

$\mu_3 = 6.6$ ، $\sigma_3 = 1.36$

ب- 4 ج- 6.2 د- $\sigma_1^2 = 2.0$ هـ- 0.97

/2 أ- الفقرة 1 ، $p = 0.90$ وهي الفقرة الأسهل.

ب- فقرة 3 $pq = 0.25$

ج- 0.50

د- 0.36

هـ- 0.22

و- 0.05

/3 أ- 9.00 12.15 7.68

25.00 18.00

16.00

ب- 125.66

/4 النتيجة ليست منطقية، لأنه مع إمكانية تحقيق التباين الأساسي فإن جدوى أو معنى الدرجات سينخفض وذلك للتقييد الشديد للمجال السلوكي الذي يمثله الاختبار.

الفصل السادس

/1 أ- منتظم

ب ـــ عشوائي

ج- منتظم

د- عشوائي

هـ- عشوائي

/2 أ- 2.5 ب- 1.5 ج- 0.5

د- 0.25 هـ- 1.25 و- 0.50

ح- لا يمكن، وذلك لأن الدرجات الحقيقية الفعلية ودرجات الخطأ غير معروفة، لذا فإن تباين كل منهما لا يمكن حسابه مباشرة.

/3 لا. مع أن $\varepsilon(E)$ = صفر لكل مفحوص عبر عدد كبير لا نهائي من الاختبارات، ومن غير المقبول أن تساوي صفر لثلاثة قياسات متكررة فقط.

/4 لا يكون لجون أكثر من درجة حقيقية واحدة، ونظرياً يوجد درجة حقيقية مختلفة لكلا الاختبارين.

/6 ب. مع أن σ_T^2 لا يمكن حسابها مباشرة من الملاحظات التجريبية، و σ_y يمكن حسابها مباشرة من قياسات في i و J على الأفراد أنفسهم.

/7 $$\rho_{yy\backslash} = \frac{\sigma_2\,\sigma_1\,\rho_{12}\;2 + \sigma_2^{\,2}\,\rho_{22} + \sigma_1^{\,2}\,\rho_{11}}{\sigma_2\,\sigma_1\;\rho_2\;\rho_1\;2 \;+ \sigma_2^{\,2} + \sigma_1^{\,2}}$$

أ- 0.55

ب- 20.93

ج- 275.97 ($\sigma_2\,\sigma_1\,\rho_2\,\rho_1\,2 + \sigma_2^{\,2} + \sigma_1^{\,2} = \sigma_y^{\,2}$)

د- 0.97 (الحل باستخدام الصيغة المبنية في تمرين 7)

هـ- 0.84 (استخدم رمز $y_1 = x_2 + x_3$

$$y_2 = x_7 + x_4 + x_1$$

$$\frac{\sigma_{37} + \sigma_{34} + \sigma_{13} + \sigma_{27} + \sigma_{24} + \sigma_{12}}{\sigma_{y2} + \sigma_{y1}}$$

و- 19.9 إلى 24.1

ز- دائري (لا علاقة).

الفصل السابع

1/ أ- اتساق داخلي، يمكن استخدام KR21

ب- صيغة بديلة أو مكافئة.

ج- معامل ألفا.

د- اتساق داخلي (يمكن استخدام KR22)

هـ- الاختبار ـــ الإعادة باستخدام الصيغة المكافئة (اتساق وتكافؤ).

2/ أ- فردي 4, 1, 4, 2, 3, 4, 2, 1, 2, 3

ب- زوجي 31, 3, 3, 2, 4, 1, 2, 2, 3

ج- 0.74 حسب التباين باستخدام ه في المقام

هـ- أقل

و- يفضل استخدام التجزئة الفردية- الزوجية.

فأول تجزئة يبدو أنها تعطي أكثر مجموعة متشابهة من الدرجات الممكن الحصول عليها من التجزئات الممكنة جميعها. لذا فإن الحساب المعتمد على هذه التجزئة لا يكون قريب جداً من معامل ألفا (أي الحد الأدنى لمعامل الدقة).

ز- 0.63 , 0.55

ط- الفقرات ليست متساوية في صعوبتها.

3/ أ- من أ : 3.3 ومن ب : 3.1

لا، يجب أن يؤخذ بعين الاعتبار حجم الخطأ المعياري للقياس بالنسبة للانحراف المعياري.

ب- 16.7 - 23.3

13.4 - 26.6

ج- 27.6 , 12.4

د- 0.83

هـ- يجعلها قيمة مغالى فيها للقيمة المحسوبة فيما لو أكمل الطلبة الاختبار.

و- 0.58 , 0.40

/4 أ- معامل إعادة الاختبار بعد يوم واحد (0.89, 0.91, 0.87, 0.95).

ب- 0.89 , 0.85

ج- 0.78 , 0.83 , 0.80 , 0.86

د- 0.68 , 0.70

هـ- كانت معاملات الإعادة بعد أسبوع لحساب معاملات التكافؤ.

الفصل الثامن

/1 أ- $\hat{\rho}_l^2 = 0.926$

ب- ∓ 15 $\quad \dfrac{\sigma e^2 + n\sigma_2^2}{n'_i} = 15 \pm 0.09$

ج- $= \hat{\rho}_{L*}^2 = 0.862$

/2 $\hat{\rho}_i = 0.355$ إذا كان للمفحوصين فرص تفضيل متماثلة، فإن الدرجة على أي فقرة يمكن تعميمها إلى درجة النطاق. علاوة على أن التعميم يظهر حتى عندما تطبق فقرات مختلفة على مفحوصين مختلفين.

/3 أ- (أ) الفقرات

(ب) بعد الفقرات من الممكن أن يكون عشوائياً، وعادة يرغب الفاحصين بالتعميم إلى مجموعات أكبر من تلك التي في الاختبار.

(ج) تباين درجة النطاق (σ_p^2) ، تباين الدرجة الملاحظة المتوقع/ n'_2

$\sigma_e^2 = \sigma_p^2$

(د) $\dfrac{\sigma_e^2 , \; MS_r - MS_p}{n_i} \quad = MS_r$

ب- (أ) البطاقات (I) السيكولوجيين (J).

(ب) على الأغلب أن كلا من البطاقات والسيكولوجيين غير ثابت.

(ج) تباين درجة النطاق $\sigma_p{}^2$ ، تباين الدرجة الملاحظة المتوقع:

$$\frac{\sigma e^2}{nj\backslash \quad ni\backslash} + nj\backslash / \sigma pj^2 + ni\backslash / \sigma pj^2$$

$$\frac{MS_r - MS_{Pj}}{ni} = \sigma^2 pj$$

$$\frac{MS_r - MS_{pj}}{n_1} = \hat{\sigma} pj^2$$

$$MS_r = \hat{\sigma}_e{}^2$$

(أ) النماذج (I) والموضوعات (J).

(ب) النماذج ثابتة.

(ج) تباين درجة النطاق $n\backslash_i / \sigma p_i{}^2 + \sigma p^2$

تباين الدرجة الملاحظة المتوقع:

$$n\backslash_j n\backslash_i / \sigma_e{}^2 + \sigma_{pj}{}^2 + n\backslash_i / \sigma_{pi}{}^2 + \sigma_p{}^2$$

$$\frac{MS_{pi} - MS_p}{n_J n_i} = \hat{\sigma}_p{}^2$$

$$\frac{MS_r - MS_{pi}}{n_i} = \hat{\sigma}_{pi}{}^2$$

$$MS_r - \hat{\sigma}_e{}^2 + \hat{\sigma}_{pj}{}^2$$

د- (أ) الفقرات (I) والتلاميذ (J).

(ب) يمكن أن يكون كلاهما عشوائي.

(ج) تباين درجة النطاق $\sigma_p{}^2$، وتباين الدرجة الملاحظة المتوقع:

$$(\sigma_e{}^2 + \sigma_{ij}{}^2 + nj\ (\sigma_j{}^2 + \sigma^2{}_{pj}) + \sigma^2{}_{pi} + \sigma^2{}_p$$

$$\frac{MS_r + Ms_{j:p} - MS_{pi} - MS_p}{n_j n_i} = \hat{\sigma}_p{}^2$$

(د)

$$\frac{MS_r - MS_{pi}}{n_j} = \sigma_{pi}^2$$

$$\frac{MS_r - MS_{j:p}}{n_1} = \sigma_j^2 + \sigma^2_{pj}$$

$$MS_r = \sigma_e^2 + \hat{\sigma}_{ij}^2$$

/4 أ- $$\frac{\sigma_p^2}{\hat{\sigma}_e^2 + \hat{\sigma}_{pj}^2 + \hat{\sigma}_{pi}^2 + \sigma_p^2} = \rho^2$$

$$\frac{27.729}{1.261 + 41.732 + 0.015 + 27.729}$$

$$0.392 =$$

ب- زيادة عدد الحوادث (المواقف الملاحظة) وذلك لأن $\hat{\sigma}_{pj}^2$ هي أكبر عنصر في تباين الخطأ لتباين الدرجة الملاحظة.

ج- تصميم 1 وبثلاثة ملاحظين، وحوادث عددها اثنان متداخلة، وتصميم 3 بثلاثة حوادث وثلاثة ملاحظين يعد تصميماً عملياً. واكبر معامل تعميم ينتج عن استخدام تصميم 2:

د- $$\frac{n_j / \sigma^2_{pj} + \hat{\sigma}^2_p}{n_i \backslash n_j \backslash / \hat{\sigma} + n_j / \hat{\sigma}_{pj}^2 + n_j / \hat{\sigma}_{pj} + \hat{\sigma}^2_p \backslash} = \hat{\rho}^2$$

$$\frac{3/41.732 + 27.729}{6/10261 + 3/\ 41.732 + 2/\ 0.15 + 27.729}$$

هـ- أكثر ملائمة للدراسة الموصوفة في د.

1/ أ- 0.029

ب- 1.69

ج- 0.036

2/ أ- 0.39

ب- لا، العلاقة من منخفضة إلى متوسطة.

ج- 0.36 0.75 0.88

د- $\frac{MS_p - MS_r}{MSp}$

هـ- 0.40

و- *KR20*

3/ أ - لدرجة القطع *0.60* *0.80*

لدرجة القطع *0.70* *0.53*

لدرجة القطع *0.80* *0.67*

لدرجة القطع *0.90* *0.87*

ب- *0.29, 0.08, 0.21, 0.42*

ج- *0.60, 0.07, 0.33, 0.73*

هـ- *0.58*

4/ أ- *0.51* (لدرجة القطع *0.50*).

ب- نظرياً *M (C)* يجب أن تكون أقل.

ج- *0.51* لدرجة القطع *0.50*

د- هي أكثر من معامل لدقة الاتساق.

هـ- قيمة المعامل يعتمد على قيمة تباين درجات النطاق

5/ أ- 0.21

6/ أ- تباين الدرجات قليل نسبياً.

ب- يمكن تحسينه.

ج- يمكن تقليله لأن ρ_1^2 يمكن خفضها.

د- قيم P^* و k أقل من قيمة P.

هـ- أقل درجة قطع.

الفصل العاشر

1/ تحديد الأهداف المهمة ونسبة الفقرات التي تقيسها، الارتباط بين تقديرات الأهمية وعدد الفقرات التي تقيسها، نسبة الأهداف التي لم تغطى بفقرات هذا الاختبار.

2/ أ-

الفقرة	هدف 1	هدف 2	هدف 3
1	<u>0.67</u>	0.08-	0.59-
2	<u>0.83</u>	0.67-	0.17
3	<u>0.50</u>	1.00-	<u>0.50</u>
4	0.50-	<u>1.00</u>	0.50-
5	0.00	0.00	0.00

ب- الفقرات التي تقيس الهدف تحتها خط في بند أ.

ج- ينخفض المعامل لأن الفقرة حكم على أنها تقيس هدفين بصورة مناسبة.

4/ كلما أصبح الاختبار أكثر تجانساً، فمن غير المحتمل أن يتضمن فقرات محتوى آخر، لذا فإن تباينات الصيغتين ستقل، وكذلك فروق الدرجات بينهما ستزداد، وينتج عن ذلك نسبة أقل.

5/ أ- يقلل التباين (يختزله).

ب- المستخدمين في الوظيفة قد لا يمثلوا مجتمع المتقدمين للوظيفة .

ج- تلوث المحك.

د- لم تجري دراسة صدق مباشرة لدرجات الاختبار مقابل المحك الذي نهتم به.

6/ أ- صدق البناء

ب- الصدق المرتبط بمحك تنبؤي.

ج- صدق محتوى.

د- الصدق المرتبط بمحك تلازمي.

هـ- صدق مرتبط بمحك تنبؤي، ومن الممكن صدق بناء.

7/ أ- لجوردن $0.66\ ,\ 0.73\ ,\ 0.64,\ 0.70$

لبيرهاريس $0.80\ ,\ 0.73\ ,\ 0.79\ ,\ 0.80$

ب- $0.80\ ,\ 0.55\ ,\ 0.59\ ,\ 0.78$

ج- المظهر الفيزيائي والملائمة الأكاديمية.

د- الارتباط بين الملائمة الشخصية والمظهر الفيزيائي كما تقاس بالطريقة أ، وبين الملائمة الأكاديمية والمظهر الفيزيائي كما تقاس بالطريقة ب.

هـ- الارتباط بين الملائمة الأكاديمية بالطريقة ب، والاجتماعية بالطريقة أ.

8/ $1.00\ ,\ 0.75\ ,\ 0.83\ ,\ 1.00$

9/ للمساعدة في الاشتقاق: يمكن تعظيم ρ_{xy} عندما تكون $1.00 = \rho_{jxjy}$

10/ تحديداً، معامل الصدق لا يجب أن يتجاوز مربع الجذر التربيعي لمعاملات الثبات. ويكون الجذر التربيعي لمعامل الثبات أكبر من المعامل عندما تكون قيمته أقل من 1.00.

11/ $\frac{(\hat{\rho}_{xx1} - 1)\ \rho {x_y}^2}{(\hat{\rho}_{xy}{}^2 - 1)\ \hat{\rho}_{xx}\backslash} \geq \leq$ (افرض أن ثبات المحك = 1). حيث أن ρ_{xy} أقل معامل صدق مرغوب.

و $\hat{\rho}_{xx\backslash}$ تقدير الثبات لمجموعة الفقرات الأولية في الدراسة. ولا يوجد ضمان بأن زيادة طول الاختبار سيزيد من الصدق إلى المستوى المطلوب.

الفصل الحادي عشر

1/ هو الارتباط بين GPA و MRA للطلبة المتجانسين في المهارات الأكاديمية.

2/ لـ $0.50 = R^2$ $0.02 = R^2_{CV}$

ولـ $0.70 = R^2$ $0.30 = R^2_{CV}$

حساب الصدق التقاطعي ضعيف عندما يحسب مربع الارتباط 0.50 أو 0.70 ويكون أكثر أهمية عندما يكون مربع الارتباط المحسوب 0.68، وهذا يقترح أن تكون النسبة 1:3 فيما لو كان معامل الارتباط للمجتمع أساسي، وبالطبع 0.85 أو أعلى. وعدا ذلك وللحصول على صدق تقاطعي مناسب يفضل النسبة الأكبر للمفحوصين بالنسبة للمتنبئ.

3/ أ- $(D. 48)\ 2.2 + (CFIT)\ 3.4 + 84.6 = y\backslash$

ب- $218.6 = y\backslash$

ج- $24.7 = \hat{\sigma}\, y.x_1.x_2$

د- 261 , 3 , 175.9

هـ - لا، حجم العينة كبير نسبياً.

1/ أ- متوسط Y للمجموعتين هو 6.59 , 8.33 على التوالي. ومتوسط الوسط 7.64. يصنف أي مريض درجته $y = 7.46$ أو أعلى منها على أنه شارب غير مضبوط.

ج- حسب القانون المذكور في أ، فإن جميع المرضى مصنفين بشكل صحيح.

د- حيث أن القانون طبق على المرضى الذين طور القانون من خلالهم، فإن دقة التصنيف تكون صحتها مؤكدة.

/1 أ- الاحتمالية المتكافئة A : $\frac{A}{B + A}$ = 0.75 لكل من الذكور والإناث.

النسبة الثابتة: $\frac{B + A}{D + A}$ = 1.14 و 0.57 للإناث (طريقة الاختيار متحيزة).

الاحتمالية المشروطة: $\frac{A}{D + A}$ = 0.86 للذكور و 0.43 للإناث (طريقة الاختيار متحيزة).

الانحدار C = -1.873 للذكور.

b = 0.055 للذكور.

C = - 2.058 للإناث

b = 0.061 للإناث (طريقة الاختيار متحيزة).

ب - Xc = 0.92 ، ذكور $\frac{A}{D + A}$ = 0.43

Xc = 0.82 ، إناث $\frac{A}{D + A}$ = 0.43

Xc = 70 للذكور و 66 للإناث.

د- لا يوجد نقطة قطع مفردة على الاختبار تتجاوز 70% من الذكور و 70% من الإناث.

/2 للمجموعة الرئيسية = 11.55، وللمجموعة الفرعية = 11.65.

/3 $E(u)_s$ = 0.80 للمجموعة الرئيسية.

$E(u)_r$ = 0.625 للمجموعة الفرعية.

$E(u)_s$ = 1.125 للمجموعة الفرعية.

$E(u)_r$ = 0.625 لذا يتم اختيار كلا المجموعتين

الفصل الثالث عشر

/1 أ- 2= ب- 0.2910

ج- خطأ في المعاينة

د- 0.378 هـ- 0.435 و- 0.357

ز- $hj^2 - \rho_{xx} = sj^2$

اللغة الاسكتلندية	0.263
اللغة الإنجليزية	0.342
التـــــاريخ	0.393

لا، كل متغير له تباين درجات حقيقي أساسي غير مرتبط بالعوامل المشتركة.

ح- هل $\hat{\rho}_{ij}$ تساوي تقريباً $a_{i1} a_{ji} + \hat{a}_{j1} + \hat{a}_{i1}$ لجميع قيم $\hat{\rho}_y$ ؟

/3 أ- لا. يوجد أكثر من عامل.

ب- لأنه يوجد أكثر من خمسة عوامل لها تشبعات أكبر من 1.0، ومن الضروري معرفة تشبع العامل لحلول العوامل 3 , 4 , 5 لرؤية ما إذا كان لها تفسيراً منطقياً. علاوة على استخدام تدوير العوامل الذي يجب أن يؤخذ بعين الاعتبار.

الفقرات من 28 إلى 31 لها تشبعات كبيرة على عاملين، وهذا يطرح تساؤلاً عما إذا كان محك البناء البسيط يحقق الحل المكتوب بالتقرير.

ج- يبدو أنه من غير المناسب القول بأن الفقرات غير مشبعة بالعامل الأول، وبالتالي الإدعاء أن الفقرة 13 لها تشبع 0.41 غير مشبعة بالعامل الأول. لاحظ تشبعات الفقرات 14 , 19 , 22 على العامل الثاني.

/1 أ- $D1 = 0.00$ للأعلى وأدنى 50%

فقرة $D2 = 0.40$

فقرة $D1 = 0.33$ لأعلي وأدنى 30%

$D = 0.67$

ب- فقرة 1 - 0.3 (حسب تباين الدرجة الكلي باستخدام N في المقام).

فقرة 2 0.66

ج- فقرة 1 -0.18

فقرة 2 0.52

د- يتطلب ارتباط بايسيريال افتراض أن المتغير التي تقيسه الفقرة توزيعه طبيعي. ومع ارتباط بايسيريال يكون التوزيع ثنائي.

/2 أ- فقرة 23

ب- عبر الذين كانت إجاباتهم صحيحة على الفقرة 25، حصل معظمهم على درجات عالية، بعكس الموقف مع الفقرة 21، إذ أن معظم الذين كانت إجاباتهم صحيحة كانوا من ذوي الدرجات المنخفضة.

ج- إعادة كتابة الاستجابة 2 كي تصبح أكثر جاذبية ومنطقية للذين ليسوا متأكدين من إجاباتهم.

د- قد يكون هنالك خللاً في الفقرة (يجب أن تكون الإجابة الصحيحة 4 وليس 3، أو أن هنالك غموضاً بالكلمات المستخدمة في نص الفقرة أو البديل 3 مما يسبب فهماً خاطئاً.

هـ- لـِ 24 0.66

لـِ 25 0.52

و- + 0.15

ز- 0.625

/3   أ- اختبار مربع كاي للاستقلالية = 33.93، لذا يوجد علاقة دالة إحصائياً.

ب- مربع كاي لعينة كبيرة ١٢,٢٥ـ لذا يوجد فروق دالة إحصائياً بين صعوبات الفقرات.

/4   أ- 3   ب- 1   ج- 2   د- 5   هـ- 4

/5   3.37 , 0.13

الفصل الخامس عشر

/1 $0.00 = \rho_{14} (\theta_2)$

$0.00 = \rho_4 (\theta)_2 \quad , \; 1.00 = \rho_1 (\theta)_2$

$\rho_1(\theta_2) \rho_4 (\theta_2) = \rho_{14} (\theta)_2$

$0.00 = \rho_{23} (\theta_2 < \theta < \theta_3)$

$1.00 = \rho_2 (\theta_2 < \theta < \theta_3)$

$0.00 = \rho_3 (\theta_2 < \theta < \theta_3)$

$\theta_3 > \theta > \theta_2$ لكل $\rho_2 (\theta) \rho_3 (\theta) = \rho_{23} (\theta)$

$0.45 = \rho (+ , + / \theta)$ /2

لكن $0.35 = 0.7 x 0.5 = \rho_1 (\theta) \rho_2 (\theta)$

$\rho_2 (\theta) \rho_1 (\theta) \quad \rho (+ , + / \theta)$ ، وبما أن الفقرات ليست مستقلة محلياً فإنه لا يكون أحادي البعد.

/3   يشير نمط الاستجابات إلى أحادية البعد، وكذلك فإن المنحنيات المميزة للفقرة، (ICC'S) دوال خطوة.

/4 $0.44 , 0.58 , 0.75 , 1.33 = ag$

$1.31- , 1.05- , 0.87-, 0.66- = bg$

المعامل P_g يعكس معاملي الصعوبة (b_g) والتمييز (a_g) لذا فإنه يفضل كدليل للصعوبة.

/5 أ- الفقرات 4 ، 7 ، 15 غير مشابهة في تمييزها لمعظم الفقرات في الاختبار (لاحظ أن معامل ارتباط بوينت بايسيريال متطرف مع درجة الاختبار الكلية).

ب- فقرة 1 تبعد كثيراً للمجموعة 4

فقرة 3 تبعد كثيراً للمجموعات 4-6

فقرة 4 تبعد كثيراً للمجموعة 1

ج- من الأسهل إلى الأصعب 1 , 4 , 3

د- 9.0 , 10.77

/6

θ	2.0-	1.5-	1.0-	0.5-	0.0	1.5	1.0	1.5	2.0
A	0.88	1.03	1.15	1.22	1.24	1.22	1.15	1.03	0.88
B	0.81	1.080	1.53	1.62	1.53	1.35	1.08	0.81	

/7 يمكن تعديل الصعوبة على مدى واسع فقط في حالة كون التمييز عالياً، ومع ذلك سيكون هنالك فقد كبير للمعلومات للمفحوصين الذين تقع درجات السمة الكامنة لهم في منطقة وسط توزيع الصعوبة.

الفصل السادس عشر

/1 dg للفقرات هي 0.64 , -0.832 , 2.121 , 1.738 , -0.537 , 0.364 , 0.385 , 0.308 , 0.310 , -0.446 والفقرات 3 و 7 تختلف قيم d لها كثيراً عن الفقرات الأخرى، لذا يبدو أنها متحيزة.

/2 $X_c^2 = 1.40$ وهذه ليست ذات دلالة إحصائية

$X_5^2 = 0.76$ وهذه ليست ذات دلالة إحصائية.

/3 باستخدام الفقرات 5-12, 13-15, 16, 17, 18, 19, 20, 21-23نحصل على تكرارات متوقعة تحقق المحك المبين في هذا الفصل. وإحصائي $X^2{}_j$ لهذه الفقرات هو 0.3 ,0.164 ,3.358 ,1.714 ,0.056 ,0.434 ,0.09 , 0.887 ومجموعها 15.507

/4 الإحصائي Z للفقرات الخمس عشر هي:

0.43 , 1.27, 0.26 , 0.10-, 1.20, 1.44, 0.55, 0.93-, 1.13, 1.20, 0.65-, 0.15-, 1.04- ومن هذه القيم إحصائي t للفقرة 6 دال عند مستوى 0.05، ومع هذا فمن المهم تذكر أن 15 فقرة تم اختبارها لمعرفة تحيزها عند مستوى 0.05 وإن كانت جميع الفقرات غير متحيزة في الحقيقة، فإن العدد المتوقع للاختبارات الدالة إحصائياً = 0.05 (10) = 0.75 وحيث أن واحداً منها دال إحصائياً، لذا فمن المناسب اعتبار الفقرات الخمس عشر غير متحيزة.

الفصل السابع عشر

1/ أ- 6,6,6,10,4

ب- المفحوص (د) لم يحذف فقرات وأجاب عن ٤ فقرات خطأ.

المفحوص (ج) حذف فقرتان وأجاب عن فقرتين خطأ.

المفحوص (هـ)حذف ٣ فقرات وأجاب عن فقرة واحدة خطأ.

2/ معامل الارتباط بين نصفي الاختبار وتصحيحه باستخدام معادلة سبيرمان-براون.

3/ ترتيب المفحوصين لا يتغير فيما لو أجاب جميع الطلبة عن جميع الفقرات.

4/ زيادة وقت الاختبار من 50 إلى 90 دقيقة مكافئ لزيادة طول الاختبار بنسبة 1.8 من طوله الأصلي. والثبات المحسوب للاختبار المطول 0.80 وهو أقل من الثبات الذي نحصل عليه بالطريقة التجريبية.

5/ أ- 5.55 لكل منهم.

ب- ج 4.75

د- 5.55

هـ- 4.95

ج- 3.94

د- 4.44

هـ 4.44

الفصل الثامن عشر

/1 27.7

/2 28.9

/3 باستخدام منهجية تعظيم القيمة الصغرى Iminimax وذلك لتعظيم أقل فقد متوقع.

Xg = 5 النسبة هي الأكثر ضرورة.

الفصل التاسع عشر

أ- استخدم الصيغة $\hat{\sigma}_M = \sqrt{\frac{\hat{\sigma}^2}{N}}$ وذلك عند N = 20، واستخدم N - 1 في المقام عند حساب قيمة $\hat{\sigma}$ (تتغير الإجابات اعتماداً على الدرجات العشرون المختارة في العينة).

ب- اختر عشوائياً أي فصلين من الفصول الخمسة. وصيغة الخطأ المعياري هي:

$\sigma\mu_c = \frac{\hat{\sigma}\mu}{\sqrt{k}}$ وفي حالة اختيار الفصلين الأول والثاني فإن :

$1.55 = \sigma\mu_c$

ج- يجب أن تتضمن العينة 4 طلاب من طلبة ESE، و 16 طالباً ليسوا من طلبة ESE.

/2 أ- منتظم ب- طبقي

ج- على مرحلتين : طبقية وعنقودية

د- معاينة عشوائية بسيطة.

/3 100

/4

الدرجة الخام X	الرتبة المئينية الخام	الدرجة الزائية الخطية	الدرجة المعيارية الزائية	الدرجة التائية	الرتبة المئينية المعيارية
45	99	2.04	2.17	9	98
44	94	1.59	1.51	8	94
43	86	1.14	1.08	7	87
42	75	0.68	0.66	6	75
41	57	0.23	0.18	5	59
40	40	0.23-	0.27-	5	41
39	26	0.68-	0.66-	4	24
38	15	1.14-	1.06-	3	13
37	7	1.59-	1.51-	2	6
36	2	2.04-	2.17-	1	2

/5 أ- كاثي كسبت أكثر في الترتيب المئيني.

ب- لا. يتضمن التساعي الرابع مدى أوسع من الرتب المئينية من التساعي الثاني.

ج- يفقد جون أكثر في الدرجات الخام.

د-لا توجد علاقة خطية بين الأنواع المختلفة من التحويلات الخطية.

/7

الدرجات الخام	الدرجات التائية	درجات الذكاءات الانحرافية	تدريج ETS
45	717	133	717
44	65.1	123	651
43	60.8	116	608
42	56.6	110	566

/1 $40.2 + (47.9 - x)\ 1.02 = y^*$

/3 فيما يلي جدول تحويل الدرجات الخام على الصيغة (ب) إلى درجات سمة كامنة على تدريج يعتمد على المفحوصين الذين تقدموا للصيغة (أ).

الدرجة الخام	درجة السمة الكامنة
1	2.06-
2	0.99-
3	0.29-
4	0.30
5	0.83
6	1.37
7	1.95
8	2.65
9	3.72

/4 الدرجات المعتمدة على درجات منفردة في الجزء الثاني يمكن تحويلها إلى تدريج الجزء الأول باستخدام المعاملة (20-1)، وبحساب قيم d, c, a إلى باستخدام المعاملات (P20، 20–ب، 20–ج)، ويمكن إضافة درجات الفقرات المفردة إلى الدرجات المعبر عنها في الجزء الأول لتكملة حل المسألة.

وبالتعويض بالمعادلات هذه، نحصل على:

$$\sqrt{\frac{(1.511 - 2.934)\ 2(1.412) + 46.22}{(4.677 - 2.934)2\ (1.881) + 36711}} = a$$

$$1.27 = \sqrt{1.606} = a$$

$$(4.300 - 4.250)\ 1.881 + 16.600 = C$$

$16.51 = A$

$17.07 = d$

/5 الخطأ المعياري للقياس هو:

$$\hat{\sigma}_e = \hat{\sigma}_x \sqrt{1 - \rho'_{xx}}$$

$$= 10\ (\sqrt{10 - 0.7})$$

$$= 10\ \sqrt{0.3}$$

ولا يكون الخطأ المعياري للمعادلة أكبر من 10/1 من الخطأ المعياري للقياس. لذا فإن الخطأ المعياري للمعادلة = 0.1 (10 $\sqrt{3}$) = $\sqrt{3}$ للتصميم B يمكن حساب قيمة N المطلوبة باستخدام المعادلة (6-20) والتي تؤدي إلى:

$$3 = \frac{2\ (100)\ (1\text{-}0.7)}{N}\ [(1+0.7)5 + 2]$$

وبحل المعادلة N = 40.